University of Alberta

"A novel mechanistic spectrum underlies Axenfeld-Rieger Syndrome associated chromosome 6p25 copy number variations."

by

Bhaskar Chanda©

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Master of Science

in

Medical Sciences-Ophthalmology

Edmonton, Alberta
Spring-2008

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-45789-4
Our file Notre référence
ISBN: 978-0-494-45789-4

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Genomic alterations (copy number variations and mutations) in *FOXC1* and *PITX2* cause Axenfeld-Rieger Syndrome (ARS). Of these, *FOXC1* is particularly interesting as it undergoes frequent copy number variations, secondary to chromosome 6p25 rearrangements. However, the underlying mechanisms of these 6p25 genomic rearrangements remain elusive. Also, the genetic basis in the majorities of ARS cases remains unknown. Hence, this thesis aims to understand the molecular mechanisms of 6p25 rearrangements and identify novel genetic factors of ARS.

In this thesis, I identified a novel spectrum of recombination, DNA repair and replication based mechanisms of ARS associated 6p25 genomic rearrangements. I also identified ARS patients with known and unknown genetic factors and made preliminary progress towards the design of a pre-symptomatic molecular diagnostic test for ARS. The study thus provides insights into the molecular mechanisms of ARS and contributes towards the better diagnosis of this frequently blinding disorder.

Acknowledgements

The work presented in this thesis has been possible due to much appreciated assistance from many people. Many thanks go to my supervisors Drs Ordan J. Lehmann and Michael A. Walter for their generous support and guidance. Furthermore, I am thankful to Dr.Yves Sauvé for his help and rounding out an excellent graduate committee.

To my labmates Mika and Ming: Thank you for making the lab a fun and organized place to work. The time spent with you guys, especially at different conferences, will be relished forever. Farideh deserves a special thank for providing many samples whenever needed. To members of the Medical Genetics: Thank you all, especially the graduate students, for not letting me feel isolated. I am also thankful to different collaborators, Drs. Jane Sowden, Nigel Carter, Andrew J Mungall, Margaret Barrow and Willam B Dobyns, who kindly provided ARS samples for my study and also for their helpful discussion.

I am also thankful to various funding agencies that generously supported my research- the Department of Ophthalmology for covering the international differential tuition and providing Alcon travel fund for conferences; the Medical Science Graduate Program for providing graduate research assistantship and the University of Alberta for Mary Louise Imrie Graduate Student Award.

Last but not the least, I would like to thank both my Parents and Shoma for their encouragement and support during the ups and downs of the tenure of my study.

Table of Contents

Subject	Page
Chapter One: General introduction	
The embryogenesis of the mammalian eye	2
The optic vesicles and optic cup	3
Anterior eye development	3
Aqueous humor dynamics	8
Glaucoma	10
Axenfeld Rieger Syndrome	11
Genetics of Axenfeld Reiger Syndrome	15
FOXC1-the forkhead box transcription factor	16
PITX2- the paired like homeo box transcription factor	21
Other genes and loci in ARS	24
Copy number variations and genomic rearrangements	25
Array comparative genome hybridization	30
Array CGH: Advantages and limitations	32
SNP arrays	37
Mechanisms of genomic rearrangements	38
Criteria for characterizing the 6p25 CNVs	45
Bibliography	48
Chapter Two: A novel mechanistic spectrum underlies glaucoma associated chromosome 6p25 copy number variations	
Introduction	67
Methods	70
Patients	70
Analysis of 6p25 copy number variants	70
Bioinformatic and haplotype analysis	73

Subject	Page
Results	75
Discussion	88
Bibliography	96
Chapter Three: Towards identifying novel genetic factors in ARS	
Introduction	103
Methods	104
PCR	104
Sequencing	105
Restriction enzyme digestion	105
Real time PCR	106
Microsatellite marker genotyping	106
Results	109
Discussion	116
Bibliography	119
Chapter Four: General discussion	
Importance of genomic rearrangements' characterization	124
Future Directions	133
Towards identifying novel genetics factors of ARS	133
Simple molecular diagnostic test of ARS	135
Bibliography	138
Appendices	
A. Montage illustrating array CGH result and genomic architecture of a 6p25 segmental deletion.	141
B. *FOXC1* and *PITX2* sequence alterations identified in ARS cases.	142

List of Tables

Tables		Page
Chapter 2		
2.1:	Ocular phenotypes of 6p25 copy number variation families	71
2.2:	Primers (5'→3') for amplifying junctional fragments, *FOXC1* qPCR and haplotype analysis	74
2.3:	Haplotype of segmental duplication pedigrees	87
Chapter 3		
3.1:	Primer pairs used for *FOXC1* and *PITX2* amplifications	108
3.2:	*FOXC1* mutations identified in previous and present studies	114
3.3:	*PITX2* mutations identified in previous and present studies	115

List of Figures

Figures		Page
Chapter 1		
1.1:	Morphogenesis of the eye	4
1.2:	Aqueous humor dynamics	9
1.3:	Ocular and non-ocular defects in ARS	13
1.4:	Fox triplet in 6p25 and FOXC1	17
1.5:	Gene and protein structure of PITX2	22
1.6:	Molecular mechanisms of genomic disorders	28
1.7:	Array comparative genome hybridization (CGH)	33
1.8:	Gene conversion and NAHR	40
1.9:	Replication based mechanism –FoSTeS	44
Chapter 2		
2.1:	Characterization of 6p25 CNVs	76

Figures **Page**

2.2: Summary of genomic architecture in 6p25 segmental duplication pedigrees (#1-6) 77

2.3: Montage illustrating the genomic architecture of the 7th segmental duplication 79

2.4: Summary of the segmental deletion's (#8) genomic architecture 81

2.5: Schematic representation of qPCR-identified segmental deletion (#9) 82

2.6: Montage illustrating the segregation of junctional fragments with disease phenotype in representative pedigrees 84

2.7A: Microsatellite marker genotyping in pedigree #1 85

2.7B: Microsatellite marker genotyping in pedigree #4 86

2.8: Mechanistic spectrum observed in the five ancestral segmental rearrangements 90

Chapter 3

3.1: *FOXC1* mutations 110

3.2: *PITX2* mutation 111

Figures **Page**

3.3: Montage illustrating the relative positions of the different microsatellite markers and the 6p25 Fox triplet 112

3.4: Montage illustrating the discrepancy between microsatellite marker genotyping and oligonucleotide array result 113

Chapter 4

4.1: Coupled NAHR and NHEJ 127

4.2: Probable mechanism of the origin of genomic architecture in pedigree #8 129

4.3: Montage illustrating the probable microsatellite markers and SNPs that can be used for haplotype analysis in pedigree #9 131

Abbreviations

AA	Axenfeld Anomaly
AC	Anterior Chamber
ACG	Angle Closure Glaucoma
ARS	Axenfeld Rieger Syndrome
AS	Axenfeld Syndrome
AnS	Angelman Syndromes
ASD	Anterior Segment Dysgenesis
BAC	Bacterial Artificial Chromosome
CB	Ciliary Body
CBE	Ciliary Body Epithelium
CE	Corneal Ectoderm
CGH	Comparative Genome Hybridization
CMT	Charcot-Marie-Tooth Disease
CN	Corneal Endothelium
CNV	Copy Number Variation
CS	Corneal Stroma
DMD	Duchen Muscular Dystrophy
DSB	Double Strand Break
E	Embryonic day
FGI	Familial Glaucoma Iridogoniodysplasia
FHD	Fork head domain
FISH	Fluorescence *In Situ* Hybridization
FoSTeS	Fork Stalling and Template Switching
HJ	Holliday Junction
HNPP	Hereditary Neuropathy with liability to Pressure Palsies
IE	Iris Epithelium
IGDA	Iridogoniodysgenesis Anomaly
IGDS	Iridogoniodysgenesis Syndrome
IH	Iris Hypoplasia

Abbreviations

IOP	Intra Ocular Pressure
IS	Iris Stroma
KCl	Potassium Chloride
LCR	Low Copy Repeat
LCRe	Low Complexity Region
LINE	Long Interspersed Nuclear Elements
LOH	Loss of Heterozygosity
LP	Lens Pit
M	Mescenchyme
MEPS	Minimal Efficient Processing Segment
$MgCl_2$	Magnesium Chloride
NAHR	Non Allelic Homologous Recombination
NCBI	National Center for Biotechnology Information
NE	Neural Ectoderm
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Signals
NR	Neural Retina
dNTP	Deoxyribonucleotide Triphosphate
OAG	Open Angle Glaucoma
OC	Optic Cup
OMIM	Online Mendelian Inheritance in Man
OV	Optic Vesicle
PAC	P1 Artificial Chromosome
PACG	Primary Angle Closure Glaucoma
POAG	Primary Open Angle Glaucoma
PMD	Pelizaeus- Merzbacher
PWS	Prader-Willi Syndromes Disease
RA	Reiger Anomaly
RE	Restriction Enzyme

Abbreviations

ReM	Medina
ROMA	Representational Oligonucleotide Microarray Analysis
RPE	Retinal Pigmented Epithelium
RS	Rieger Syndrome
SC	Schlemm's Canal
SE	Surface Ectoderm
SD	Standard Deviation
SDSA	Synthesis Dependent Strand Annealing
SINE	Short Interspersed Nuclear Elements
SMS	Smith Magenis Syndromes
SNP	Single Nucleotide Polymorphism
TM	Trabecular Meshwork
UCSC	University of California–Santa Cruz
UPD	Uniparental Disomy
VCFS	Velocardiofacial Syndrome
WBS	Williams-Beuren Syndrome

Chapter 1:

General Introduction

Introduction

"Equipped with his five senses, man explores the universe around him and calls the adventure Science." ~Edwin Powell Hubble, *The Nature of Science*, 1954. Of these five senses, vision contributes the most to our learning and memory since 70% of the human body's sense receptors are clustered in the eye [1]. However, blindness resulting from ocular disorders has been ever growing. Therefore, research to prevent blindness by understanding the molecular basis of frequent blinding disorders, and subsequently developing a therapy to combat the same, is a highly noteworthy goal. Following the above goal, the research described in this thesis aims to understand the molecular mechanism of a blinding disorder- Axenfeld–Rieger Syndrome (ARS). In the following sections, a comprehensive review of ocular development, the genetic basis of Axenfeld-Rieger Syndrome, copy number variations and molecular biology techniques relevant to this thesis is presented.

The embryogenesis of the mammalian eye

Ocular disorders including ARS can result from developmental defects in ocular structures during the morphogenesis of the eye. Understanding the eye morphogenesis will serve as the entry point to understanding developmental ocular disorders.

The optic vesicles and optic cup

The development of the mammalian eye begins with the appearance of the optic vesicles as lateral outgrowths of the prosenchephalon (around embryonic day (E) 8.5 in mouse and about four weeks of gestation in human) [2]. When the outgrowths contact the surface ectoderm, they stimulate the formation of the lens placode (Figure 1.1A) [2-4]. The lens placode forms an enlarged trench in the ectoderm, the lens pit, which transforms into the lens vesicle and is held to the lens ectoderm by the lens stalk. Simultaneously, a double layered optic cup is formed when the distal part of the optic vesicle is invaginated into its more proximal part (Figure 1.1B). The inner layer of the optic cup gives rise to the neural retina, while the outer layer differentiates into the retinal pigmented epithelium. Once both the optic cup and the lens vesicle have formed, the later detaches from the surface ectoderm. The surface ectoderm subsequently plays an important role in the anterior eye development [2-4].

Anterior eye development

The anterior segment of the eye is primarily formed by the differentiation of three types of tissues: surface ectoderm, neural ectoderm and periocular mesenchyme. At the time of the differentiation of these tissues, the migration of the mesenchymal cells occurs in three waves from the neural crest.

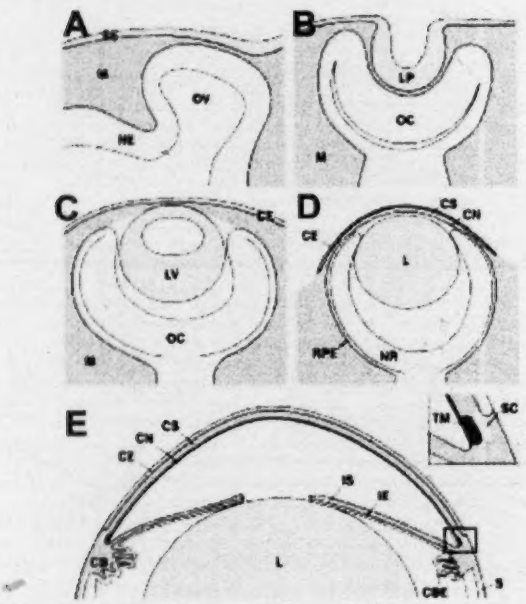
Figure 1.1: Morphogenesis of the eye.

Top panel: Schematic representation of the use of different tissues during the formation of the eye, taken from [4].
A. Movement of the optic vesicles brings the neural ectoderm in contact with the surface ectoderm.
B. Formation of the lens pit from the surface ectoderm and invagination of the optic vesicles to form the optic cup.
C. The surface epithelium gives rise to the corneal epithelium while the lens vesicle detaches from the future cornea.
D. The corneal endothelium and corneal stroma are formed from the migrated mesenchymal cells. The formation of the lens is completed by this time.
E. Formation of the iris separates the fluid filled cavity surrounding the lens into the anterior and posterior chamber.

Bottom panel: Formation of the trabecular meshwork, taken from [2]
F. Accumulation of mesenchymal cells (solid arrows) occurs at the idiocorneal angle.
G. Mesenchymal cells (solid arrows) become flattened and get separated by extracellular fibers, while vessels appear adjacent to the sclera (open arrow).
H. The extracellular fibers form the trabecular beams and the sclera vessels give rise to the Schlemm's canal.

Abbreviations:

SE: Surface ectoderm
M : Mescenchyme
NE: Neural ectoderm
OV: Optic vesicle
LP: Lens pit
OC: Optic cup
CE: Corneal ectoderm
TM: Trabecular meshwork
AC: Anterior Chamber
ReM: Medina
CN: Corneal endothelium
CS: Corneal stroma
NR: Neural retina
RPE: Retinal pigmented epithelium
IE: Iris epithelium
IS: Iris stroma
CB: Cilliary body
SC: Schlemm's cannal
CBE: Ciliary body epithelium


A
B
LP
OC
C
LV
OC
D
CS
CN
L
RPE
E
TM
SC
CS
CN
CZ
IS
OE
L
CBE
F
AC
CB
G
AC
CB
H
AC
CB

(1) First wave of migration: Following the detachment of the lens from the surface ectoderm, the periocular mesenchymal cells migrate between the anterior epithelium of the lens vesicles and the surface ectoderm (Figure 1.1C) [5] and is observed at E12 in mouse and around the sixth week of gestation in human [2]. As the number of mesenchymal cells increase they form several layers that are separated by extracellular matrix. In parallel, the lens fibre fills the lens vesicle and the cavity within the lens disappears. Following this, the surface ectoderm gives rise to the corneal epithelium while the mesenchyme layer gives rise to the corneal endothelium. The mesenchymal cells between the corneal epithelium and the endothelium differentiate into corneal stroma fibroblasts which are responsible for the synthesis of the unique extracellur matrix that provides corneal transparency [2, 6-9].

(2) Second wave of migration: During the corneal endothelium differentiation, the lens detaches from the future cornea. The differentiation of the corneal endothelium gives rise to the rudimentary Descemet's membrane. At this time, the second wave of migration of mesenchymal cells occurs from the neural crest and these cells differentiate into the corneal stroma. The Descemet's membrane thus lies between the corneal endothelium and the corneal stroma [2].

(3) Third wave of migration: The detachment of the lens generates a fluid filled cavity between the anterior surface of the lens and the cornea. Following this, the neural ectoderm of the optic cup migrates into this cavity along the anterior surface of the lens to form the epithelia of the iris and the pupillary sphincter muscles [3, 10]. Concurrently, the ciliary body epithelia are also formed. The mesenchymal cells from the neural crest migrate along these newly formed epithelial layers to form the stroma of the iris and the ciliary body. Thus, the iris and ciliary body have both ectodermal (neural ectoderm) and mesodermal (mesenchymal cells) origin. A fully developed iris will have different layers (from back to front): pigmented epithelia; pupillary sphincter muscles; vascularized stroma; and an anterior layer of chromataphores and melanocytes. The formation of the iris separates the fluid filled cavity between the lens and cornea into the anterior and posterior chambers (Figure 1.1D) [2, 9, 11]. As will be discussed in the later part of this thesis, incomplete development of the iris, also known as iris hypoplasia (IH), is often associated with ARS and glaucoma.

The last structures to develop during anterior eye development are the trabecular meshwork and Schlemm's canal [12, 13]. At E 17-19 in mouse or the 15^{th}-20^{th} week of gestation in human, a thick mass of mesenchymal cells appears at the iridocorneal angle which remain separated by extracellular fibres (Figure 1.1F). These extracellular fibres self-assemble into a lamellar structure that remains covered by endothelial-like cells. At the end of this process, a spongy tissue called the

trabecular meshwork (TM) is formed (Figure 1.1.E-G). The junction between the TM and corneal endothelial is characterized by the presence of the Schwalbe's line - the termination of the Descemet's membrane. Congenital opacity of Schwalbe's line, which then appears as a white line in slit-lamp examinations, is one of the common diagnostic criteria of ARS [14].

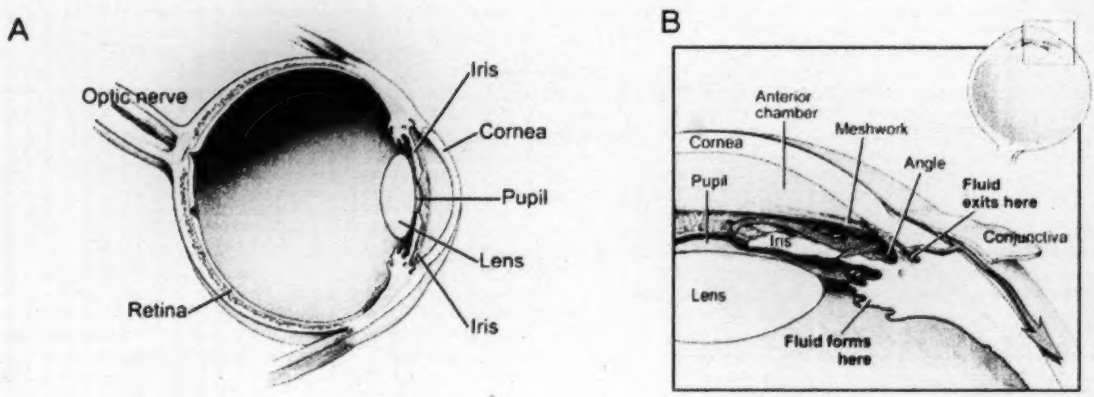
The Schlemm's canal (SC), which connects to the outside of the TM, originates from the coalition of sclera vessels which lie next to the idiocorneal angle (Figure 1.1H). Some mesenchymal cells, by establishing contact with the trabecular beams on one side and endothelial lining of the SC on the other side, give rise to the juxtacanalicular or cribiform layer of the TM. It is this portion of the TM that offers the most resistance to the aqueous humor outflow [2].

Aqueous humor dynamics

Intraocular pressure (IOP) homeostasis is essential for the maintenance of the normal structure of the eye and to keep proper distances between the retina and refractive surfaces of the cornea and lens. This homeostasis is maintained by the continuous production and drainage of the aqueous humor. The rate of aqueous humor production by the epithelium of the ciliary body is approximately 1.5 to 2μl per minute [15]. The aqueous humor passes through the posterior chamber and the pupil into the anterior chamber where it subsequently exits through the TM. The TM serves as the filtering matrix of the aqueous humor before its pressure-dependent

Figure 1.2. Aqueous humor dynamics

A. Vertical cross-section of the mature human eye, taken from NEI (Catalog number NEA08).
B. The flow of aqueous humor, taken from NEI (Catalog number NEA11).
The path of aqueous humor flow, from the posterior chamber to the anterior chamber and subsequent drainage through the trabecular meshwork (located at idiocorneal angle) is indicated with the blue arrow



drainage into the Schlemm's canal (Figure 1.2) [2]. As such, malformation of the TM can lead to increased resistance to the aqueous humor outflow, resulting in elevation of IOP. This elevated IOP, is often, though not always, associated with glaucoma.

As evident form the above discussion, developmental defects of the ocular anterior segment-anterior segment dysgenesis (ASD)-can lead to ocular disorders including ARS and developmental glaucoma [4].

Glaucoma

Glaucoma is one of the leading causes of 'irreversible' blindness worldwide [16-18]. It represents a heterogeneous group of optic neuropathies with a complex genetic basis. Glaucoma is characterized by glaucomatous cupping or gradual loss of retinal ganglion cells that leads to progressive narrowing of the visual field and eventual blindness, if untreated. The loss of retinal ganglion cells is often associated with increased intraocular pressure [(IOP)>21mmHg]. Elevated intraocular pressure generally results from increased resistance to the drainage of the aqueous humor. Elevated intraocular pressure and glaucomatous cupping are recognized as the hallmarks of glaucoma, however, these symptoms may not be present in a portion of glaucoma cases.

Glaucoma can be classified based on [19]

a) Etiology (primary versus secondary)

b) Anterior chamber anatomy (open angle versus closed angle)

c) Chronology (juvenile versus adult)

Primary open angle glaucoma (POAG) is characterized by elevated intraocular pressure that results from increased resistance to the aqueous humor outflow primarily at the juxta-canalicular portion of the trabecular meshwork. However, the angular width of the anterior chamber could remain normal. In contrast, narrowing of the angular width is the characteristic feature of primary angle closure glaucoma (PACG). It is estimated that there will be 60.5 million people with open angle glaucoma (OAG) and angle closure glaucoma (ACG) in 2010, increasing to 79.6 million by 2020, and of these, 74% will have OAG [20].

Axenfeld-Rieger Syndrome

Of the many ocular disorders resulting from ASD and leading to developmental glaucoma, Axenfeld-Reiger Syndrome (ARS) is particularly interesting as 50% of ARS patients develop glaucoma [21, 22]. The Axenfeld-Rieger anomalies were first reported by Vossius [23] and Darwin [24] and subsequently characterized as a genetic disorder by Axenfeld [25] and Reiger [26]. The Axenfeld–Rieger group of disorders was previously diagnosed differentially, as Rieger syndrome (RS), Reiger

anomaly (RA), Axenfeld anomaly (AA), iridogoniodysgenesis syndrome (IGDS), iridogoniodysgenesis anomaly (IGDA), familial glaucoma iridogoniodysplasia (FGI), and iris hypoplasia (IH) [27]. The typical clinical features of the subgroups are described below. From this, the reasons behind the adoption of the present widely accepted, unified title of Axenfeld–Rieger syndrome (ARS) will become evident.

Rieger syndrome or RS is characterized by ocular, dental and abdominal defects (Figure 1.3) [28]. One of the major ocular anomalies associated with RS, posterior embryotoxon, is characterized by a prominent Schwalbe's line in slit-lamp examinations as a white or yellowish ring lining the peripheral cornea [29]. There are also several cases of RS where, instead of posterior embryotoxon, the angle tissue becomes abnormal, and often include iridocorneal adhesions and hypoplastic iris [18, 30, 31]. Other symptoms include polycoria (appearance of multiple pupils due to holes in the iris), corectopia (displaced pupils), tearing eyes, thick/cloudy cornea, megalocornea (large cornea) and buphthalmos (large eye). These later four features, may indicate elevated IOP, the major risk factor associated with glaucoma development in the ARS patients [27].

Although not of direct relevance to this thesis, dental and abdominal anomalies are important features of RS and will be described briefly. Dental anomalies encompass a spectrum of anodontia (lack of teeth), hypodontia (fewer teeth) and microdontia

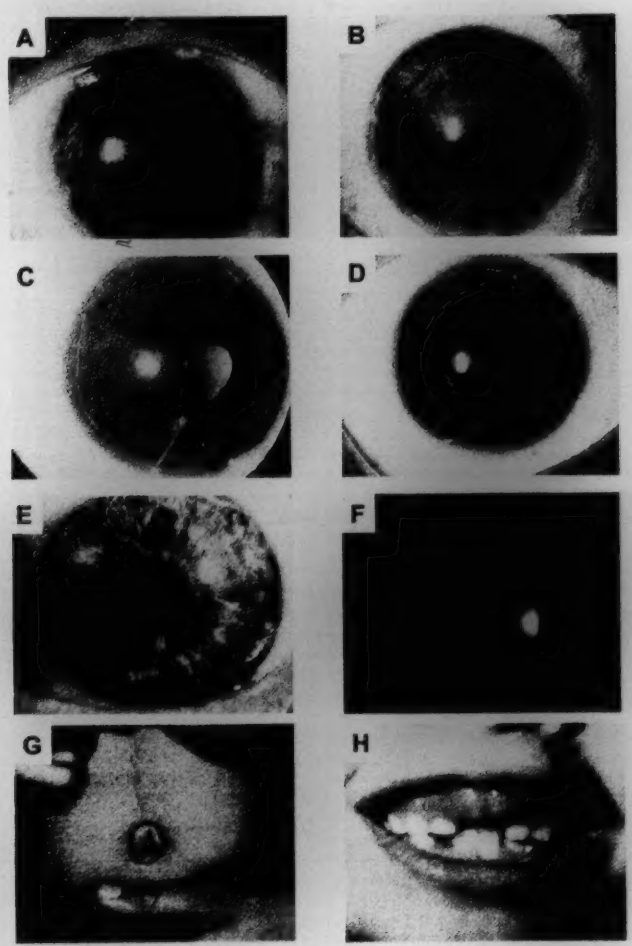
Figure 1.3 Ocular and non-ocular defects in ARS

(**A-F**) Some of the ocular defects found in ARS. A-E taken from [27] and F taken from [32]

A. Displaced pupil and Schwalbe's line, polycoria.
B. Abnormal pupil and pseudopolycoria.
C. ASD with posterior embryotoxon.
D. Anteriorly displaced Schwalbe's line.
E. Congenital ectropion of the iris.
F. Buphthalmos.

(**G-H**) Some of the non-ocular features found in ARS, taken from [33]

G. Periumbilical skin is present in excess
H. Dental abnormalities



(smaller teeth) [34]. The final hallmark of RS relates to abdominal defects, the most common of which is redundant and often hyperplastic periumbilical skin. While low severity symptoms include abnormal protrusion of the umbilical stump, extreme cases include death of affected patients resulting from omphalocele - failure of the abdominal wall to close [27]. Gut defects include, anteriorly misplaced or imperforate anus. Other defects associated with RS include empty sella syndrome and growth hormone deficiencies resulting in growth retardation [35, 36], cardiac anomalies [37, 38], hearing loss [38], and mental retardation [39, 40].

RA is characterized with the same ocular defects as RS, but without the nonocular defects, while AA describes patients with glaucoma that have angle tissue defects and posterior embryotoxon, but no systemic features [41]. IGDA is characterized by hypoplastic iris and glaucoma [42], but without posterior embryotoxon and systemic features [43]. However, IGDA shows genetic linkage to chromosome 6p25 [44], one of the ARS loci, and is now considered a part of the Axenfeld-Reiger group of disorders. IGDS shows similar features of IGDA, with addition of some systemic features found in RS. Later on, studies have shown that mutation in *PITX2,* an ARS causing gene, is associated with IGDS [45]. A detailed description of the 6p25 locus and the *PITX2* gene are provided in the later part of this thesis. Iris hypoplasia (IH), characterized by the presence of hypoplastic, discoloured iris and the absence of posterior embryotoxon, iridocorneal adhesion,

corectopia and polycoria, was also initially described as a separate syndrome [46]. However, as can be found in Online Mendelian Inheritance in Man (OMIM), IH is considered to be the same phenotype as IGDS [41, 47, 48] and in addition, *PITX2* mutations have been shown to cause IH [49]. For these reasons, IH is now considered under the central umbrella of ARS.

As evident from the above discussion, the Axenfeld-Reiger group of disorders displays a wide range of clinical manifestations and associated intensities, resulting in a strong overlap of the various symptoms. As such, their sub-classification have little value and as Alward stated in 2000, "*While these disorders can be split into small groups because of minor differences, it seems more logical to combine them under the umbrella of Axenfeld-Rieger syndrome*" [41].

Genetics of ARS

ARS is a rare disorder where the genetic transmission has mainly been found to be autosomal dominant. However, there are occasional reports of other modes of inheritance, as in noninheritable AA [50], or autosomal recessive RS [51]. The genetic heterogeneity of ARS is apparent from the fact that two genes (*PITX2* and *FOXC1*) and three other chromosome loci are associated with this disorder.

***FOXC1*- the forkhead box transcription factor.**

Forkhead genes have been identified in a wide spectrum of species, and their essential roles in developmental and cellular processes, including tumorigenesis, cell cycle control [52-60], and most recently in immunity development [61-65], have been studied in depth. Fox genes, or Forkhead box genes, are characterized by the presence of forkhead DNA-binding domain (FHD)-a conserved 110 amino acid residues [66]. Chordate Fox genes are classified into 17 clades (A to Q) based on the amino acid sequence of their forkhead domains [67]. X-ray crystallography revealed the three-dimensional structure of an HNF-3/forkhead DNA-recognition motif complexed with DNA, which is composed of three alpha helices, anti-parallel beta strands and random coils (Figure 1.4). Whilst alpha helix 3 interacts with the major groove of the DNA, the random coils and the beta strand are known to interact with the minor groove. The arrangement of the random coils gives rise to a butterfly-like appearance, and because of this forkhead box transcription factors are also known as "winged-helix" transcription factors [68, 69].

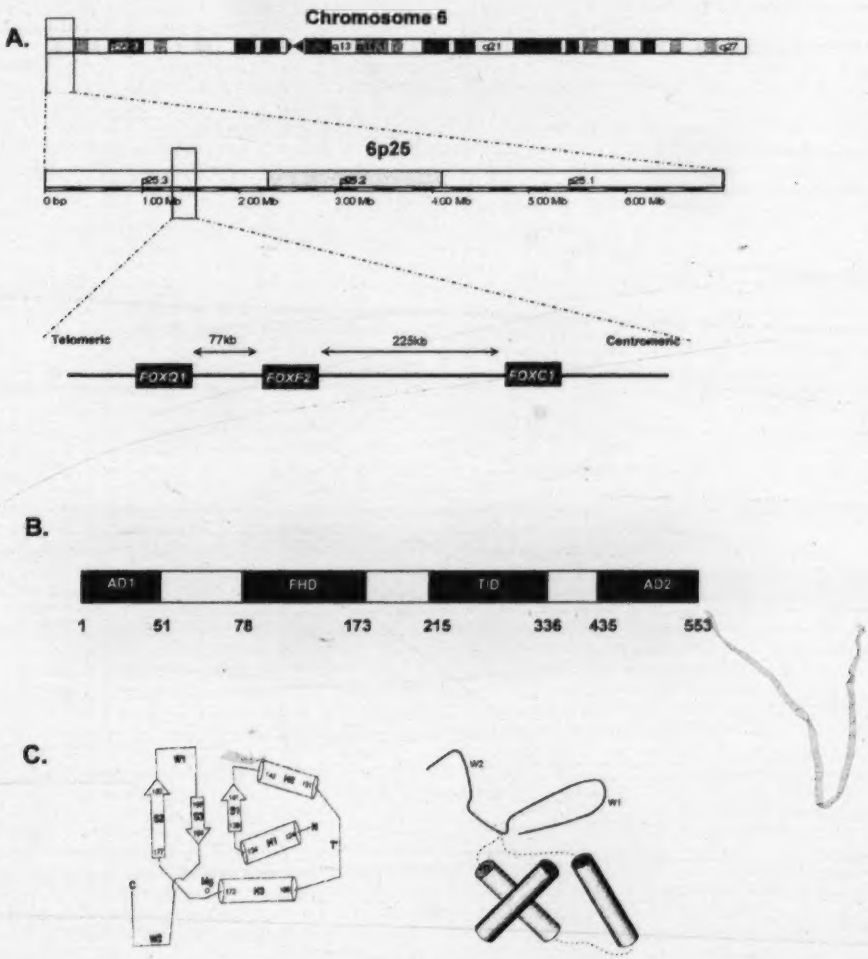
The number of forkhead genes increases from invertebrates to vertebrates, primarily by gene duplications. This idea is supported by the combined presence of the same combinations of clades on different chromosomes [D3/E3 (1P32) D4/E1 (9q22)], common functions of close paralogues [C1(6P25)/ C2(16q24)] and the existence of similar clusters [6p25: FOX(*Q1/F2/C1*)) and 16q24 [*L1/F1/C2*]] [69].

Figure 1.4 Fox triplet in 6p25 and FOXC1

A. The genetic localization of the 6p25 Fox triplet (*FOXQ1*, *FOXF2* and *FOXC1*) and relative distances between them.

B. Domains of FOXC1. The green rectangles indicate transactivation domains, while the red and black rectangles represent forkhead and transcriptional inhibition domains respectively.

C. Secondary structure [68] and winged helix appearance [69] of the forkhead domain (FHD).



The forkhead box genes can exist in a triplet cluster [6p25: *FOX(F2/Q1/C1)*], duplet [1p32: *FOX(D3/E3)*] and singleton [7p31: *FOXP1*] on different chromosomes [69]. Interestingly, of the several Fox genes known to have definite ocular phenotypes (*FOXC1, FOXC2, FOXL2, FOXE3* etc.), *FOXC1* on 6p25 is one of the earliest and best studied genes- in both human and murine models.

The importance of 6p25, with respect to ocular disorders, was highlighted in 1996 when the genetic localization of iridogoniodysgenesis anomaly (IGDA) was described [44]. In the next year, Axenfeld-Reiger anomaly (ARA) [70] and familial glaucoma iridogoniodysplasia (FGI) [71] were mapped to 6p25. As noted in the description of ARS, the above mentioned ocular disorders now come under the central umbrella of ARS. All these studies heightened the importance of 6p25 and it was assigned the standard nomenclature *IRID1* (MIM: 601631). In 1995, 6p25 was shown to be the locus of the human forkhead gene *FOXC1* [72] and three years later, two separate groups independently identified genomic alterations in *FOXC1* to underlie ARS [73, 74]. Interestingly, in one of these studies, two families with ocular abnormalities (ASD) were linked to 6p25 but genetic analysis and mutational screening excluded *FOXC1* as the disease causing gene [73]. This study supported the existence of a second genetic factor (gene/ potential regulator) in 6p25 to underlie ocular abnormalities [73].

The interest in 6p25 further gained ground when altered dosage of *FOXC1*, secondary to 6p25 rearrangements (duplication), was hypothesized to cause ocular abnormalities [75]. However, the possibility of second glaucoma causing gene in 6p25 undermined this hypothesis [73]. The bona fide support to this hypothesis came from the work of Nishimura *et al.* [76] and Lehmann *et al.* [77] when they independently showed that chromosomal duplications, encompassing *FOXC1,* can result in ARS. In 2002, the evidence of 6p25 deletion [78] (encompassing *FOXC1*) causing ARS further supported this hypothesis. In addition, biochemical experiments have confirmed that less than 50% or more than 150% of normal activity of *FOXC1* causes human disease [79]. Interestingly, functional interactions between FOXC1 and PITX2 have been shown to underlie the sensitivity to *FOXC1* gene dose in ARS [80].

FOXC1 is a single exon gene, while the encoded protein- FOXC1- has several domains. The FOXC1 fork head domain (FHD), encompassing residues 69- 178, has a consensus binding sequence of RTAAAYA (Figure 1.4). It has been shown that binding of the isolated FHD produces a ninety degree bend in the DNA double helix [66]. Following the discovery of *FOXC1* mutations in ARS cases, Saleem *et al.* [81] showed that missense mutations in *FOXC1*-FHD can results in reduced transactivation and/or altered DNA binding specificity of the FOXC1 protein. In 2002, Berry *et al.* [82] identified other functional elements of FOXC1. Two essential nuclear localization signals (NLS) were identified in FHD of FOXC1: NLS1

(residues 78-93) and NLS2 (residues 168-176). Of these, only NLS2 is sufficient for complete localization of FOXC1 within the nucleus. Also, the existence of two NLS further highlights the importance of FHD, other than binding to DNA. Berry *et al.* [82] also identified two transcriptional activation domains at the N- and C-terminal regions of FOXC1: AD1 (residues 1-50) and AD-2 (residues 436-553), respectively (Figure1.4). Furthermore, GAL4 assay revealed that AD2 can mediate a greater degree of transactivation than AD1. A transcriptional inhibitory domain (residues 215-366) was also found in FOXC1 [82] and was shown to attenuate the transactivation of AD1 and/or AD2. Later, it has been shown that this transcriptional inhibition domain is the target of mitogen-activated protein (MAP) kinase, which in turn regulates FOXC1 stability and transcriptional activity [83].

Histological analysis revealed a wide range of tissues that can harbor *FOXC1* expression. *FOXC1* transcript was observed in the optic nerve head, ciliary body, trabecular meshwork and RPE. In the mouse eye, *Foxc1* is primarily expressed in the periocular mesenchyme while by E16.5, *Foxc1* expression is restricted to the region of the future trabecular meshwork [7]. In addition, *Foxc1* expression has also been observed in the iris and cornea. Based on these expression data, it has been suggested that *FOXC1* gene expression in the trabecular meshwork and ciliary body might regulate genes essential for their structure and function, and thus regulates the IOP [84].

The *FOXC1* gene has also been well characterized in the murine model. In 1998, Kume *et al.* first demonstrated that the classical mouse mutant phenotype *congenital hydrocephalus* is attributable to the *Foxc1* mutation Q123X [85]. The *Foxc1*-knockout mice ($Foxc1^{+/-}$) developed defects in the anterior segment of the eye including iris malformation, idiocorneal adhesion and corneal opacity [4]. In addition, $Foxc1^{+/-}$ mice also manifest cardiac, renal and skeletal abnormalities [85-88]. $Foxc1^{-/-}$ mice develops more severe defects than $Foxc1^{+/-}$ mice. Multiple congenital abnormalities including hydrocephalus and anterior segment dysgenesis are characteristic features of $Foxc1^{-/-}$ mice [85, 89]. In these mice, anterior segment malformation includes disorganized stroma, thickened epithelium and failure of lens to separate from cornea [7]. Very recently, mutations in *Foxc1* have been shown to cause cortical dysplasia and skull defects in mice [90]. This combination of data from patients with *FOXC1* alterations and altered/modified *Foxc1* mice suggest the importance of *FOXC1* in normal ocular development.

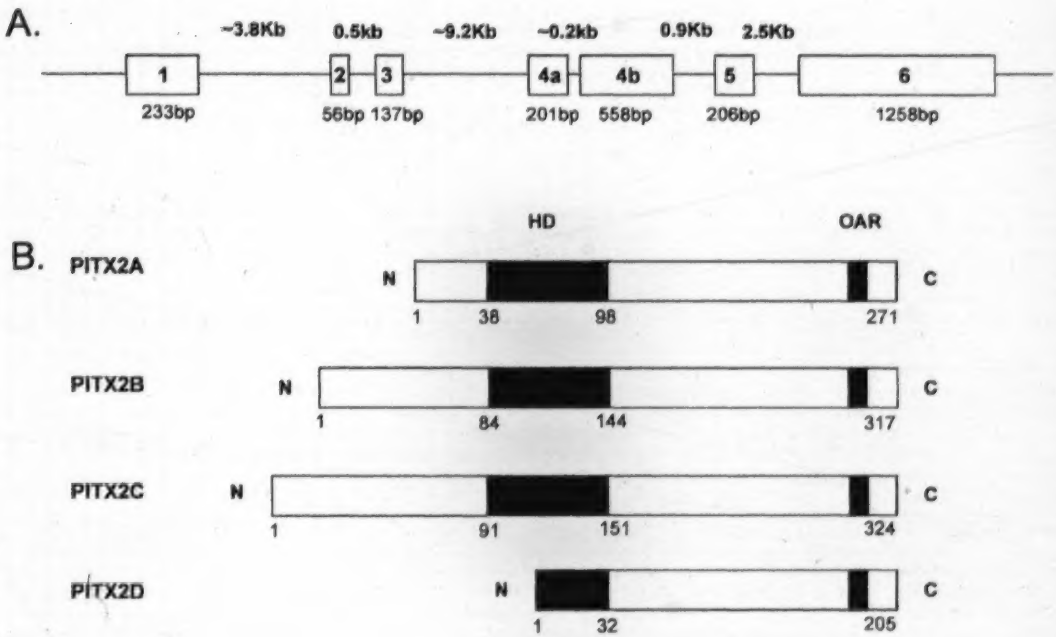
PITX2 - the paired-like homeobox transcription factor

The first ARS locus was identified to be 4q25 through combined use of cytology and linkage analysis [46, 91-93]. Subsequently, a *paired-like* homeobox transcription factor, *PITX2*, was identified by positional cloning and confirmed though mutational screening as one of the causative genes of ARS [94].

Figure 1.5 Gene and protein structure of *PITX2*

A. The introns and exons of *PITX2* and the relative spacing between them. The exons are depicted with rectangles and their sizes are shown below the rectangles. Taken from [84].

B. Different isoforms of PITX2. The homeodomain (HD) is represented with a red rectangle and the brown rectangle denotes '**o**tp, **a**ristaless, and **r**ax' or **OAR** domain. Adopted from [84].



The homeobox domain was simultaneously and independently identified by two groups of researchers as a conserved DNA sequence in the homeotic genes of the drosophila Antennapedia and Bithorax complexes [95, 96]. Later, this conserved sequence was shown to encode 60 amino acid residues [97]. The homeobos transcription factor *PITX2* is a multi-exon gene and three major isoforms, *PITX2*A, *PITX2*B and *PITX2*C, have been identified to date [94, 98-100] (Figure 1.5). The only distinguishing feature between these isoforms is the dissimilar N-terminal domains. A fourth minor isoform, *PITX2*D, was identified from a human craniofacial library [101]. Splicing of exon 4a to a cryptic 3' splice site in exon 5 produces *PITX2*D. This isoform of *PITX2*, has a truncated homeodomain and a complete C-terminal tail, and as expected, the lack of a functional homeodomain prevents this isoform from binding to DNA [101].

The three major *PITX2* isoforms have been shown to differentially regulate organogenesis. However, the molecular mechanism for this developmental preference of the *PITX2* isoforms remained unknown [84]. The alternative splicing of *PITX2* is believed to provide a mechanism for fine-tuning of gene expression during development. A significant amount of research has been done to understand the functional role of *PITX2*, by elucidating both upstream [102-110] and downstream [111-118] targets of the PITX2 protein.

Both mutations and copy number variations in *PITX2* have been shown to cause ARS. It has been proposed that the *PITX2* mutations resulting in less than 70% of normal PITX2 activity or more than 150% of normal PITX2 activity can cause human disease and that residual PITX2 activity correlates directly with the severity of anterior segment dysgenesis phenotypes [79]. In murine models, gene targeting of *Pitx2* has generated null, hypomorphic and conditional (floxed) alleles [100, 119]. Similar to human *PITX2* mutations, *Pitx2*$^{+/-}$ mice exhibit pleiotropic effects- a single gene influencing multiple phenotypic traits. *Pitx2*$^{+/-}$ mice exhibit both ocular and non-ocular defects. The ocular defects include AR-like full-thickness iris tears and irregular pupillae, cataract and microphthalmia while maloccluded incisors and reduced growth of the body are some of the non-ocular features. Homozygous null or Pitx2$^{-/-}$ mice exhibit more severe ocular phenotypes (absence of extraocular muscles, persistent lens pit, optic nerve coloboma, etc.) and systemic defects (incomplete neural tube closure, failed heart separation etc.) than Pitx2$^{+/-}$ mice [100, 119]. However, Pitx2 null mice are not of much clinical relevance as patients with PITX2$^{-/-}$ have not been reported, probably due to the lethality of the PITX2$^{-/-}$ phenotype.

Other genes and loci in ARS:

In addition to the above mentioned genes, chromosome 13q14 has been implicated in ARS from the study of cases with deletions in this region [120] and linkage analysis [121], but the causative gene in this locus has not been identified yet.

Another locus, 16q24, has been implicated in ARS [122], where the transcription factor *MAF* has been suggested to be a strong candidate gene [33, 123]. Also, a single incidence of *PAX6* (11p13) deletion has been identified in an ARS patient [124].

As evident from the above discussion, the two predominate genes associated with ARS are *FOXC1* and *PITX2*. Precise dosage of both is critical for the normal eye and other organs' development. Of these two genes, *FOXC1* is especially interesting because it undergoes frequent copy number alterations or copy number variations (CNVs), secondary to chromosome 6p25 rearrangements.

Copy Number Variations and Genomic Rearrangements:

The era of molecular medicine began with the recognition of sickle-cell anemia as a molecular disease. In the year 1949, for the first time, Linus Pauling's seminal work - "*Sickle Cell Anemia, a Molecular Disease*", demonstrated that an abnormal protein (hemoglobin) could cause a human disease (sickle cell anemia) [125, 126]. Following this, in 1956 Vernon Ingram demonstrated - "*A Specific Chemical Difference between Globins of Normal and Sickle-cell Anemia Hemoglobins*" [127] and a year later, his finding - "*Gene Mutations in Human Hemoglobin: The Chemical Difference between Normal and Sickle Hemoglobin*" [128]- represented

the first classical proof of gene mutations as the cause of a human disease. Since then, for more than five decades, considerable amount of research has been done to answer the question: how do mutations alter DNA and consequently affect the expression, structure and function of the encoded protein? However, following the completion of the human genome project, scientists have started to acknowledge that mechanisms for some genetic diseases are best understood at the structural level.

Variation of the human genome occurs at multiple levels, from sequence to structural. Sequence variation refers to SNPs and mutations, whilst structural variations, also known as copy number variations (CNVs), encompass deletions, insertions, duplications and large-scale copy number variants [129, 130]. One of the first evidences of chromosome copy number variations influencing a human phenotype came from the study of the genomic disorder Down's syndrome where an additional copy of chromosome 21 was observed [131]. Since then the number of genomic disorders, defined as disorders that result from structural changes of the genome due to certain architectural features that render a portion of the genome unstable, has been ever mounting [132-134].

Deletion and duplication genomic rearrangements lead to structural alterations of DNA segment [135]. As such, genomic rearrangements can be considered as the 'cause' and copy number variations the 'effect'. The boost to study the potential contributions of CNVs with respect to human variations was provided by two

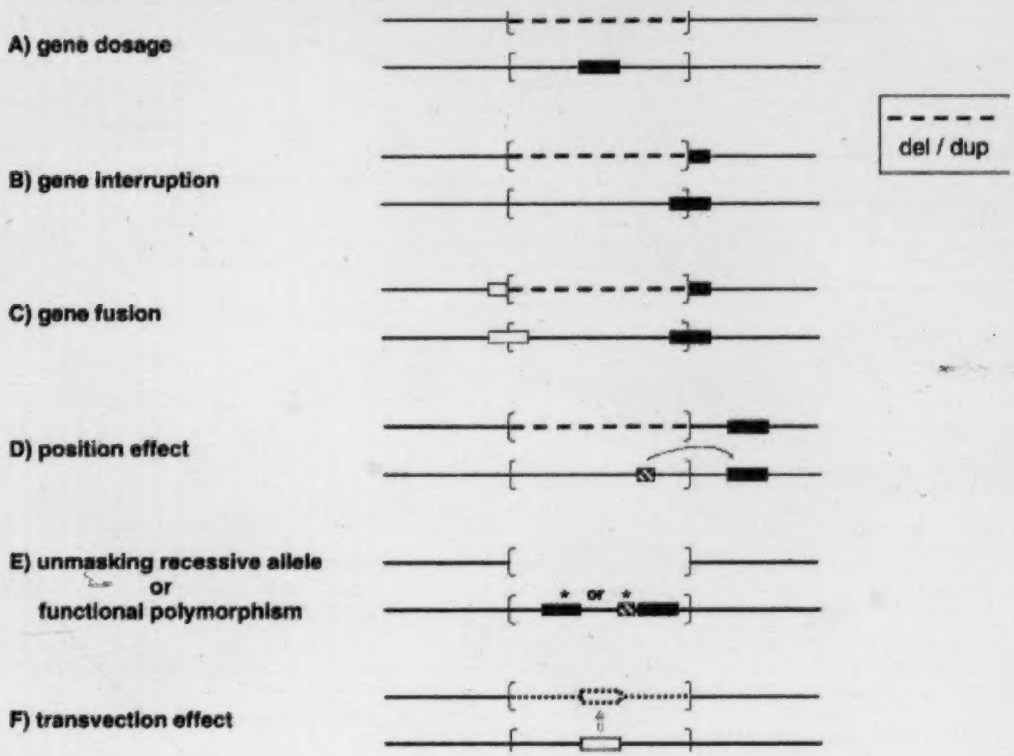
classical papers: "*Detection of large-scale variation in the human genome*"- Irafate *et al.* [136] and "*Large-scale copy number polymorphism in the human genome*"- Sebate *et al.* [137]. Irafate *et al.* identified 255 copy number variable regions in 55 normal individuals; Sebate *et al.* identified 76 variable regions in 20 normal individuals. These studies were further confirmed and extended by many other studies [138-141]. Moreover, the importance of CNVs has been strengthened by the revelation that 18% of gene expression traits can be influenced by or are associated with CNVs [142]. At this point, it should be noted that CNVs can lead to genomic disorders via different ways: dosage alteration, gene interruption, gene fusion, position effect, transvection effect and functional polymorphism (Figure 1.6) [143]. In dosage alteration, the gene lies within the rearrangement, while in gene fusion and gene interruption, the breakpoints of the rearrangements lie within the genes (Figure 1.6.A-C). In position effect, the rearrangement alters potential regulatory elements of a gene and thereby alters the expression or activity of the gene (Figure 1.6.D) [130]. Transvection effect or gene conversion is mediated when the deletion of a gene or its surrounding elements affect the communication between the alleles, while in functional polymorphism or unmasking recessive allele, the deletion of a dominant allele results in hemizygous expression of a recessive mutation (Figure 1.6.E-F).

The first initial step towards a comprehensive database of CNV came from several pioneering studies that undertook genome wide surveys of CNVs [136-139, 141, 144-147].

Figure 1.6 Molecular mechanisms of genomic disorders

The rearranged genomic interval is depicted by brackets. The dashed lines corresponds to either deleted or duplicated genomic regions, while an absent line indicates deletion with phenotypic effects from the remaining allele unmasked because of the rearrangement, and a dotted line represents deletion but where phenotypic effects result from the absence of interactions between alleles (i.e., transvection effects). The solid black rectangle represents the gene of concern, while regulatory region is shown as a hatch-marked rectangle. The asterisks denote point mutations. Taken from [143].

A. Gene dosage; **B.** Gene interruption; **C.** Gene fusion; **D.** Position effect; **E.** Unmasking recessive allele or functional polymorphism; **F.** Transvection effect.



The database, combining the results of these studies, estimates that up to 12% of our genome is structurally variant and this in turn makes copy number variation a greater cause of human genetic variation than single nucleotide polymorphism (SNP) [148, 149]. However, it should be noted that SNP databases developed through a series of phases during which researchers were able to accurately genotype, validate (or invalidate) and characterize the properties of a large number of SNPs. The rapid growth in SNP detection technologies and extensive effort to calibrate and standardize these detection methods, to maximize sensitivity and minimize false positives, further added to the accuracy of the SNP database. As such, researchers rely on these databases with confidence. Furthermore, the announcement of the '1000 Genomes Project' that aims to find rarer SNPs occuring at 1% frequency, is expected to make the SNP database more accurate than ever [150]. In contrast, the present CNV database is derived from several studies, each using a different technology platform and data processing algorithms, on less than 1000 normal individuals' genome [151]. Also, the experimental standardization and validation varies amongst these studies [151]. In order to rely on the current CNV databases more confidently, it is important that the structural variation in the general population be accurately catalogued. This is particularly essential to discriminate the pathogenic CNVs from the nonpathogenic ones. Coupled with these, various limitations of current CNV-detection technologies, as discussed in the later part of this thesis, have contributed towards the infallibility of the present CNV databases. We are thus in a situation where, even after having large CNV databases for the

normal human genome, the true estimate of CNVs' prevalence remain incompletely deciphered. The words of the great Einstein, "*Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted*", perfectly describe the situation.

Although structural or copy number variations, have long been recognized, only recently has the magnitude of their contribution been appreciated. The estimations of the prevalence of CNVs and association studies on a genome wide scale have been made possible due to the recent advancement in CNV detection technologies. Comparative Genome Hybridization (CGH) and Single Nucleotide Polymorphism (SNP) arrays have been extensively used for identifying CNVs on a genome wide scale.

Array Comparative Genome Hybridization

Comparative genome hybridization (CGH), a type of DNA micro-array, is one of the main methods of analyzing and identifying CNVs [152]. It was first developed with the goal to rapidly map DNA amplifications in tumors [153] . Initially, this was achieved using fluorescence *in situ* hybridization (FISH) onto control metaphase spreads. One of the major drawbacks incurred from the usage of metaphase FISH was the low resolution (5-10 Mb) [153]. This problem was potentially solved with the advancement of the Human Genome Project, which provided large – insert clone

libraries to replace metaphase chromosome for CGH [154-156]. In 1997, matrix CGH [157] was reported and subsequently, in 1998 array CGH [158] was performed with arrays of clones that were accurately mapped on to the human genome and spotted robotically onto glass slides [152]. In array CGH, the test and reference DNAs are differentially fluorescently (generally Cy3 or Cy5) labeled before hybridization. The resulting fluorescent ratio of each clone is the yard stick for estimating the variation of a particular region of the genome corresponding to that clone. It is worth mentioning here that while the resolution of array CGH is a function of the number, distribution, and length of the probes, it is the signal-to-noise ratio and probe response characteristics that determine the accuracy in detection of copy number variations.

The interpretation of array CGH results is quite simple. The resultant intensity or the ratio of the cohybridization of the differentially labeled probes (test to reference genome) is generally represented in logarithmic scale ($\log_2$) along the 'Y'-axis, while the position of the chromosome corresponding to the probes is represented along the 'X'-axis (Figure 1.7). Ideally, for a probe, $\log_2=0$ corresponds to a null CNV, while $\log_2<0$ indicates a deletion and $\log_2>0$ indicates a duplication of the genomic region represented by the probe. Sometimes, depending on the extent and spacing of the probes, instead of the individual probes the average $\log_2$ value of continuous probes is of more importance in identifying CNVs. For instance, in PAC/BAC array CGH individual probe's (~150kb) $\log_2$ values are considered, while

in oligonucleotide array CGH average $\log_2$ value of contiguous probes (~75-mer) is generally more useful in identifying CNVs (Figure 1.7).

Since the initial description of array-CGH almost a decade ago, several improvements have been made, most notable amongst which is the increased resolution at which CNVs can be detected. The major contributing factors for this increase in resolution of array-CGH were increasing the probe density and decreasing the size of the probes themselves [152]. Array-CGH, based on the DNA sequences, can be classified into the followings:

1) Clone arrays: This can be subclassified according to the length of the clones.
    a) large insert clones (40-200kb)
    b) small insert clones (1.5 – 4.5 kb),
    c) cDNA clones (0.5 – 2kb).
2) PCR-product arrays (100bp – 1.5kb).
3) Oligonucleotides array (25-80bp).

Array CGH: Advantages and limitations

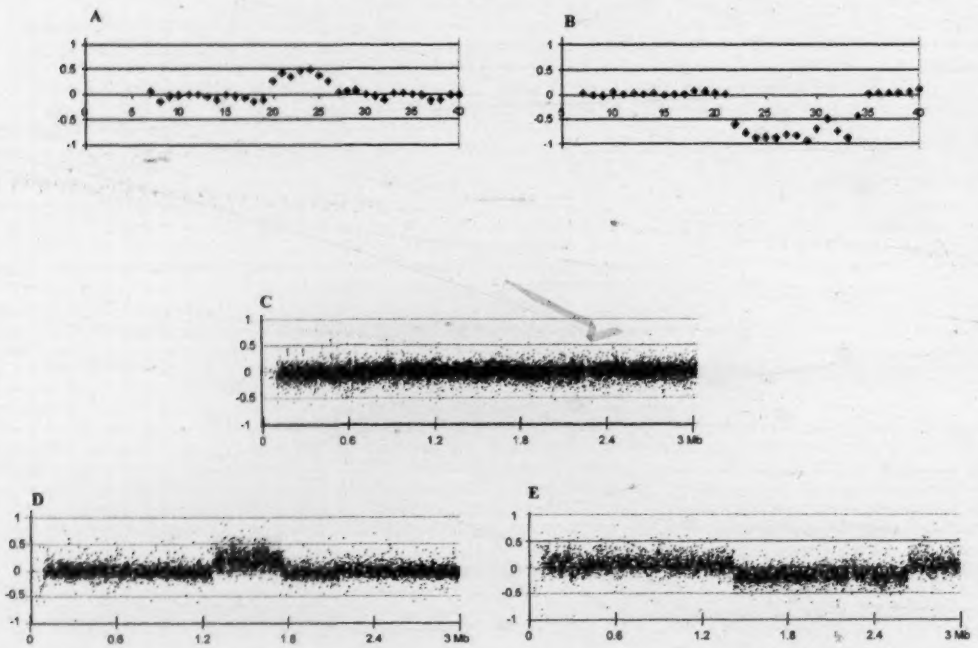
Advantages

One of the main advantages of array-CGH lies in the co-hybridization of the test and reference DNAs which reduces the influence of probe concentration and variation in slide production and processing on the final result (the ratio).

Figure 1.7 Array comparative genome hybridization (CGH)

(**A-B**) BAC-array CGH on chromosome 6p25. The extent of individual probes is ~ 150-200kb. (Photo courtesy Dr. Nigel P.Carter)
A. The probes with $\log_2$ ratio >0 represent duplication of the corresponding genomic region.
B. The probes with $\log_2$ ratio <0 represent deletion of the corresponding genomic region.

(**C-E**) Oligonucleotide array CGH on chromosome 6p25. The extent of individual probe is ~75mer.
C: Representation of a normal genomic region with null CNV.
D. Representation of genomic region experiencing duplication (shaded with light blue).
E. Representation of a genomic region experiencing deletion (shaded with light blue).



With ever increasing resolution of array CGH, the screening of genomic regions for putative CNVs is becoming more efficient. Oligonucleotide array CGH (HD2 array from NimbleGen® - 2.1 million probes) provides one of the highest resolutions (~1.5kb per array) in identifying CNVs. Such resolution is at least 10^2 times more than FISH that uses PAC/BAC as probes and 10^3 times more than cytogenetic analysis using microscopy. Such increased resolution of array CGH has made the estimation of the widespread variation of submicroscopic CNVs possible [146, 147, 152].

Another advantage of oligonucleotide array CGH is the flexibility of the probe design. In such customized arrays, by synthesizing overlapping oligonucleotides with a single basepair shift, it is possible to achieve almost nucleotide level resolution for the area of interest. This method could be very useful for custom interrogation of a specific region. Using this approach one study was able to map the translocation breakpoints within 4bp [159]. As can be found in Chapter 2 of this thesis, we have used customized array CGH (NimbleGen®) to define the breakpoints of 6p25 rearrangements. NimbleGen's® flexible Maskless Array Synthesis technology enables interrogation of a particular region of the genome at desirable resolution. The maximum average resolution of our customized array was 100bp, which proved helpful in defining the breakpoints of 6p25 rearrangements. In addition to the flexibility of the probe design, another advantage of the NimbleGen®'s customized oligonucleotide array CGH is the isothermal (Tm=76°C)

design of the probes. Probe lengths are adjusted (45mer - 85mer) to perform equivalently at a given stringency, such as AT- and GC-rich regions, in the genome. Such isothermal array design enables uniform probe performance and helps to minimize hybridization artifacts and/or bias.

Limitations

No technology is perfect, and array CGH is no exception. Despite all the promising advantages it has, it still possesses certain limitations. One of its greatest limitations is the difficulty in detecting certain types of rearrangements (such as balanced translocation and inversion). Since balanced translocation and inversion, not associated with gain or loss of genomic DNA, does not produce much alterations in the signal ratio of fluorescently labeled test and reference DNA, they remain difficult to be identified by array CGH.

At present, CNVs smaller than 500bp can be potentially detected by sequence analysis while CNVs, at a resolution >1.5kb, can be identified by a single HD2 array (NimbleGen®). As such, in terms of resolution, the present challenge is to identify CNVs between 500bp to 1.5 kb in a genome wide scale. To achieve this ~500bp resolution, a three fold increase in the probe density (~6 million probes) for a single array is required [152]. Alternatively, three HD2 arrays (NimbleGen®) can be used simultaneously, with a three fold increase in the cost, to achieve such resolution for genome wide interrogation of CNVs. The above problem can be partially

circumscribed by using customized array CGH, where probe intervals can be adjusted with respect to the gene density in the human genome.

To increase the resolution of array CGH, the probe density has been increased with time. However, an increase in probe density results in a poor signal-to-noise ratio. This drawback of increased probe density can be typically observed in oligonucleotide array CGH where the standard deviation (s.d.) of $\log_2$ ratio (0.25) is much higher than the s.d. of BAC clone-based array (0.05). In order to ameliorate this problem, an approach to reduce the complexity of the genomic DNA is sometimes used. This technology, called representational oligonucleotide microarray analysis (ROMA), digests the genomic DNA using a restriction enzyme into small fragments which are amplified with universal primers before hybridization. Unfortunately, ROMA also imposes its own drawbacks in CNV detection. Two primary drawbacks are noted by Dr. Nigel Carter [152], (i) False CNV detection due to differential representation of certain parts of the genome and (ii) probe ratio dependence on restriction fragment size and thus individual digestion pattern.

Finally, the other limitation of the oligonucleotide array CGH is the criteria used to design the probe sequence. The oligonucleotide probes are generally designed by considering only the repeat masked fraction (approximately 55%) of the human genome. As such, repeat prone regions, comprising low copy number repeats and

segmental duplications, are not interrogated by oligonucleotide array CGH [152]. However, this limitation can be overcome by the use of customized oligonucleotide arrays that provide flexible and maskless probe design, encompassing the genomic regions of interest.

SNP arrays

SNP array, essentially another DNA microarray, was initially developed to genotype SNPs, but subsequently has been adopted for identifying copy number variations. In contrast to CGH, the hybridization of oligonucleotide probes to test genomes only is carried out for CNV identification [160]. In SNP arrays, PCR based reduction of genomic complexity is used prior to hybridization of test genomic DNAs to oligonucleotide probes. The reduced complexity of genomic representation is achieved by restriction digestion of the genomic DNA, followed by ligation of the digested fragments to adaptors prior to their PCR based amplifications [161].

One of the main advantages of the SNP array is its ability to detect both physical copy number variations and genetic aberrations such as loss of heterozygosity (LOH), uniparental disomy (UPD), etc [160, 161]. The sensitivity and precision of the SNP array can be increased by averaging the signals from several probes. The current Affymetrix® Genome-Wide Human SNP Array 6.0 has ~ 946,000 probes for the detection of CNVs. As such, an effective resolution of ~3.3kb genome wide

can be achieved. However, since the SNP-probes are not uniformly distributed, the resolution of the SNP array varies across the genome. The other disadvantage of the SNP array comes from the modification of genomic DNA, which is used to reduce its complexity prior to hybridization [152]. Such modification increases the possibility of biased amplification of different regions of the genome and in turn, detection of false CNVs.

Different caveats are associated with CGH and SNP arrays, but they still provide a robust platform to identify CNVs. With ever increasing resolution in both, true estimate of copy number variations in human are soon to follow. The current estimates of CNVs prevalence (~12% genome wide) and evidence of CNV-associated several genomic disorders, suggest that CNVs underlie a substantial portion of human genetic variation and disease. Therefore, elucidating the mechanisms that underlie the origin of these CNVs is central to our understanding of human evolution and disease, and is thus warranted.

The mechanisms of genomic rearrangements

Both genetic and environmental factors contribute to genomic rearrangements. However, analyzing the contribution of environmental factors lies outside the scope of this thesis. Certain architectural features of a particular region of the human genome can render it unstable, thereby leading to genomic rearrangements. Such a

genetic factor has long been implicated to cause genomic rearrangements but still remained incompletely defined. Based on the current understanding, the mechanism of genomic rearrangements can be broadly classified into two categories.

a) Recombination based and b) Replication based.

a) Recombination based mechanisms

Based on the nature of substrates that can mediate genomic rearrangements, recombination based mechanisms of genomic rearrangements can be subclassified into two groups: i) Non Allelic Homologous Recombination (NAHR) and ii) Non-Homologous End Joining (NHEJ)

i) NAHR: NAHR, regarded as the dominant mechanism for genomic disorders, is attributable to recurrent and some nonrecurrent genomic rearrangements. Current understanding favors a double strand break (DSB) repair model, or a synthesis dependent strand annealing (SDSA) process, or double Holliday junction (HJ) dissolution as the process of NAHR (Figure 1.8) [132, 161]. In general, NAHR is mediated by nonallelic homologous DNA sequence or paralogues and is often associated with gene conversion. NAHR has been implicated in several genomic disorders including Charcot-Marie-Tooth Disease (CMT) [162]; Hereditary Neuropathy with liability to pressure palsies (HNPP) [163]; Smith Magenis

Figure 1.8 Gene conversion and NAHR

Top panel: Mechanisms of gene conversion, taken from [161].

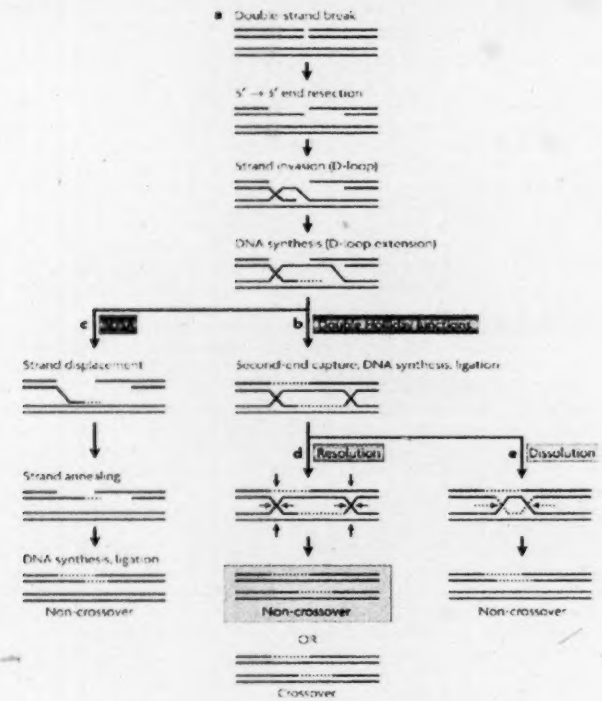
A→B→D The double-strand break (DSB) repair process

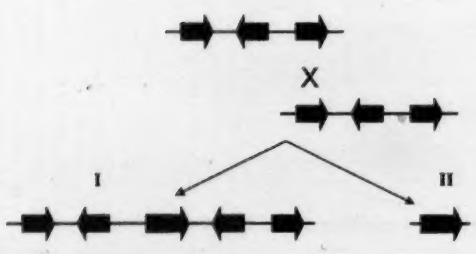
A→C Synthesis dependent strand annealing (SDSAa)

A→B→E Double Holliday junction (HJ)

Bottom panel: Schematic representation of a typical NAHR, adopted from [164].

A typical NAHR, between paralogous DNA segments (green and blue filled arrows), can give rise to both duplication (I) and deletion (II).





syndromes (SMS) [165]; velocardiofacial syndrome (VCFS) [166, 167]; Williams-Beuren Syndrome (WBS) [168]; and Prader-Willi/Angelman syndromes (PWS/AnS) [169].

The use of substrates in NAHR varies from region specific low copy repeats (LCRs or segmental duplications) to genome wide nonspecific *Alu*-repetitive elements [143, 164, 170, 171]. Although the LCRs are generally 10 - 500 Kb in size and greater than 95% identical [132, 164], the minimal homology requirement for homologous recombination, also known as minimal efficient processing segment (MEPS), is ~200–300 bp of uninterrupted homology [164]. LCRs are proposed to originate from segmental duplications of the genome and are generally comprised of genes, pseudogenes, gene fragments, repeat clusters and other chromosomal segments [171]. They are unevenly distributed, with pericentromeric and subtelomeric regions of the genome having the highest cluster [171]. The genome wide frequency of the LCR is ~ 5-10% [172]. LCRs appear to be associated with *Alu* elements that are identified at the junctions of genes/pseudogenes [173]. Coupled with this, it has been proposed that the evolution of the LCRs and *Alu* enrichment occurred at the same time [174]. As such, it is hypothesized that *Alu* enrichment (35 to 40 million years ago) sensitized the ancestral genome for *Alu-Alu-* mediated recombination events which might have initiated the expansion of LCRs - predisposing the human genome to structural rearrangements via NAHR [171].

ii) NHEJ: The term "non-homologous end joining" or NHEJ was coined in 1996 by Moore and Haber [175]. Both NHEJ and NAHR are the major processes by which double strand breaks (DSB) are repaired in mammals. However, in contrast to NAHR, the genetic factors behind NHEJ are poorly understood. The essential concept of NHEJ is that broken ends of DNA are brought together, irrespective of the sequence homology, and then they are ligated. Biochemical studies have revealed that Ku-complex is responsible for bringing the broken ends together [176]. NHEJ is an error prone mechanism that uses limited sequence homology or none at all [177, 178]. A hallmark of NHEJ is the misinsertion or deletion of nucleotides at the junction of rearrangements [177, 179-183]. Mainly found to mediate nonrecurrent rearrangements, NHEJ has been implicated in many genomic disorders including Duchenne Muscular dystrophy (DMD) [184]; Pelizaeus- Merzbacher disease (PMD) [185, 186]; deletion 1p36 [187]; and duplication 10q24 [188].

b) Replication based mechanism:

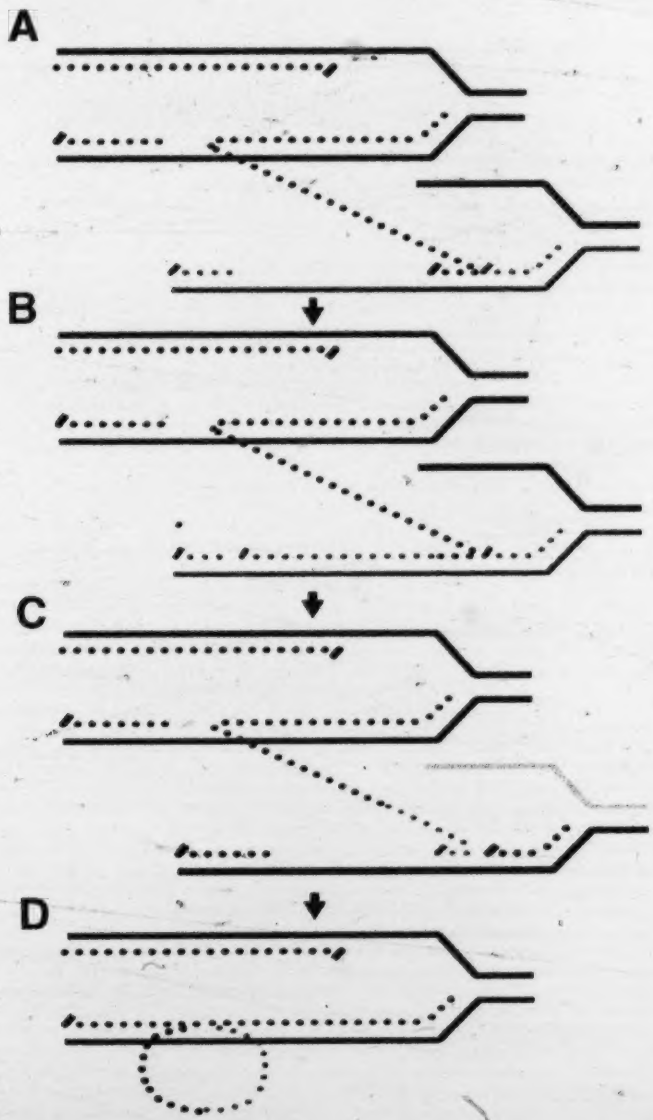
DNA undergoes replication during cell division – a key step to sustain life and development. However, this process of replication is not error proof. It is estimated that multicellular organisms with a genome of some 10^9 residues have a replication error rate of up to 10^{-9} at the nucleotide level [189]. A key feature of replication is the replication fork. The replication fork ensures that both strands (leading and lagging) of DNA are replicated simultaneously. It is worthy mentioning here that

although replication of the leading strand of DNA is continuous, the lagging strand is synthesized in a discontinuous fashion. At first, small fragments of DNA, known as okazaki fragments, are synthesized in the 5'→3' direction, and then these DNA fragments are ligated–resulting in a continual lagging strand replication.

During the process of replication different natural impediments [190, 191], such as the presence of GAG [186]-the DNA polymerase pause site, can stall the replication fork. Interestingly, these sites of fork stalling have been shown to experience elevated DNA recombination [192]. In 2008, replication fork stalling has been implicated in mediating a new mechanism of genomic rearrangements - **Fo**rk **S**talling and **Te**mplate **S**witching (**FoSTeS**) [193].

Characterization of two Pelizaeus-Merbacher disease (PMD)-associated X-chromosomal duplications revealed unique genomic architectures that were not attributable to previously known mechanisms of genomic rearrangements-NAHR and NHEJ [193]. To address the origin of these unique rearrangements, the FoSTeS model of genomic rearrangement was proposed. This model, based on a model suggested for E. *coli* gene amplification, assumes that during the process of replication, fork stalling and microhomology results in template switching of the lagging strand to find another active replication fork (Figure 1.9A). The advancement of this active replication fork could be in any direction. Once the lagging strand finds an active replication fork the sequence will be copied from the

Figure 1.9 Replication based mechanism –FoSTeS
(**A-D**) FoSTeS model, taken from [193].
A. The replication fork (dark blue and red solid lines) stalls due to natural impediments and the micro-sequence homology promotes the template switching (solid red to solid green) of the replication fork.
B. DNA synthesis (dotted green line) occurs with the movement of the replication fork in the new template.
C. The original fork disengages (dark blue and red solid lines) and may invade another replication fork (grey and black solid lines).
D. Finally the original replication fork may switch back to its original template (red solid line).



second template (Figure 1.9B). The replication fork, if again halted by natural impediment, can invade another active replication fork and sequence may be copied from the third template (Figure 1.9C). The above process can undergo several repetitions before the stalled replication fork proceeds normally (Figure 1.9D) [193].

Criteria for Characterizing the 6p25 CNVs

Chromosomal rearrangements of 6p have long been implicated to cause autism [194], craniofacial malformations, hypotonia and hydrocephalus, and renal abnormalities [195-197]. However, it was the genetic localization of iridogoniodysgenesis anomaly (IGDA) [44], Axenfeld-Reiger anomaly (ARA) [70] and familial glaucoma iridogoniodysplasia (FGI) [71] that highlighted the importance of 6p25 with respect to ocular disorders. In 1998, the congenital hydrocephalus (*ch*) mouse was shown to result from mutation in *Mf1*or mouse *Foxc1* [85]. The same study also showed that *Foxc1* mutant mice exhibit ocular defects and that 6p25 rearrangements, encompassing human *FOXC1*, are associated with congenital abnormalities [85]. In the same year, two groups independently showed that genomic alterations of *FOXC1* result in ARS [73, 74].

In 1999, the altered dosage of *FOXC1*, secondary to 6p25 rearrangements (duplications), was hypothesized to cause ocular abnormalities [75]. This hypothesis was further supported when 6p25 segmental duplications and deletions,

encompassing *FOXC1*, were shown to cause ARS [76-78]. Since then, several 6p25 rearrangements have been identified and characterized with different techniques, such as microsatellite marker genotyping [198], FISH [199, 200] and DNA microarray [199, 201] to understand their underlying mechanisms. However, the absence of base pair level characterization in the above studies precluded the analysis of the causation of the identified 6p25 rearrangements. The discovery of 6p25 segmental duplications and deletions made it one of the rare regions of the human genome to experience both deletions and duplications. Interestingly, in some rare incidences, 6p25 rearrangements excluding *FOXC1* have been reported to be associated with ocular disorders [73, 202, 203]. These data suggest that 6p25 may harbor putative genes or potential regulators of crucial genes whose genomic alterations (dosage alterations and/or position effect) could lead to oculo-developmental defects.

The importance of 6p25 is further highlighted by the presence of the Fox genes triplet cluster (*FOXQ1, FOXF2, FOXC1*) [69]. The importance of *FOXC1* in normal ocular and other organs' development has been discussed previously. The functions of the other two Fox genes (*FOXF2* and *FOXQ1*) have mainly been gathered from studies in the murine model. Although to date no human phenotypes have been associated with these genes, the role of *Foxf2* in gut development of mouse [204] and *Foxq1* in differentiation of hair in satin mice, along with evidence for *Foxf2* expression in different tissues including central the nervous system and the eye of

mice [205], suggest that these Fox genes could play a vital role in human development. In spite of all the above factors that make 6p25 an important and very interesting region of the human genome, the mechanisms that mediate 6p25 rearrangements remain elusive. Determining the causation of ARS associated 6p25 rearrangements is expected to provide insights into the molecular mechanism of genomic rearrangements, and by implication, ARS.

Bibliography:

1. Wire, M., *Sight's Impact On Your Senses: Why It Makes Sense To Care For Your Vision.* 2002.

2. Cvekl, A. and E.R. Tamm, *Anterior eye development and ocular mesenchyme: new insights from mouse models and human diseases.* Bioessays, 2004. **26**(4): p. 374-86.

3. Barishak, Y.R., *Embryology of the eye and its adnexae.* Dev Ophthalmol, 1992. **24**: p. 1-142.

4. Gould, D.B., R.S. Smith, and S.W. John, *Anterior segment development relevant to glaucoma.* Int J Dev Biol, 2004. **48**(8-9): p. 1015-29.

5. Haustein, J., *On the ultrastructure of the developing and adult mouse corneal stroma.* Anat Embryol (Berl), 1983. **168**(2): p. 291-305.

6. Dublin, I., *[Comparative embryologic studies of the early development of the cornea and the pupillary membrane in reptiles, birds and mammals].* Acta Anat (Basel), 1970. **76**(3): p. 381-408.

7. Kidson, S.H., et al., *The forkhead/winged-helix gene, Mf1, is necessary for the normal development of the cornea and formation of the anterior chamber in the mouse eye.* Dev Biol, 1999. **211**(2): p. 306-22.

8. Pei, Y.F. and J.A. Rhodin, *The prenatal development of the mouse eye.* Anat Rec, 1970. **168**(1): p. 105-25.

9. Reneker, L.W., et al., *Formation of corneal endothelium is essential for anterior segment development - a transgenic mouse model of anterior segment dysgenesis.* Development, 2000. **127**(3): p. 533-42.

10. Mann, I., *The development of the human eye.* 1964, New York: Grune & Stratton, Inc..

11. Flugel-Koch, C., et al., *Disruption of anterior segment development by TGF-beta1 overexpression in the eyes of transgenic mice.* Dev Dyn, 2002. **225**(2): p. 111-25.

12. Reme, C. and S.L. d'Epinay, *Periods of development of the normal human chamber angle.* Doc Ophthalmol, 1981. **51**(3): p. 241-68.

13. Smith, R.S., et al., *The mouse anterior chamber angle and trabecular meshwork develop without cell death.* BMC Dev Biol, 2001. **1**: p. 3.

14. Schields, M.B., *Textbook of Glaucoma.* 1997, Philadelphia: Williams and Wilkins.

15. Armstrong, R.A. and S.N. Smith, *The genetics of glaucoma.* . Optometry Today, 2002. **41**: p. 30-33.

16. Aung, T., et al., *Prevalence of optineurin sequence variants in adult primary open angle glaucoma: implications for diagnostic testing.* J Med Genet, 2003. **40**(8): p. e101.

17. Thylefors, B., et al., *Global data on blindness.* Bull World Health Organ, 1995. **73**(1): p. 115-21.

18. Quigley, H.A., *Number of people with glaucoma worldwide.* Br J Ophthalmol, 1996. **80**(5): p. 389-93.

19. Lehmann, O.J., *The Role of forkhead Genes in Eye Development and Disease*, in *Molecular genetics*. 2002, University College London: London.

20. Quigley, H.A. and A.T. Broman, *The number of people with glaucoma worldwide in 2010 and 2020.* Br J Ophthalmol, 2006. **90**(3): p. 262-7.

21. Shields, M.B., *Axenfeld-Rieger syndrome: a theory of mechanism and distinctions from the iridocorneal endothelial syndrome.* Trans Am Ophthalmol Soc, 1983. **81**: p. 736-84.

22. Shields, M.B., et al., *Axenfeld-Rieger syndrome. A spectrum of developmental disorders.* Surv Ophthalmol, 1985. **29**(6): p. 387-409.

23. Vossius, A., *Kongenitale anomalien der iris.* Klin Mbl Augenheilk, 1883(21): p. 233-237.

24. C, D., *Animals and plants under domestication 1.* 1893, D Appleton & Co.: New York. p. 434-461.

25. T, A., *Embryotoxon Corneae posteris.* Ber Dtsh Ophthalmol Ges, 1920. **42**: p. 381-382.

26. H, R., *Dysgenesis mesodermalis corneal et iridis.* Z Augenheilk, 1935. **86**: p. 333.

27. Hjalt, T.A. and E.V. Semina, *Current molecular understanding of Axenfeld-Rieger syndrome.* Expert Rev Mol Med, 2005. **7**(25): p. 1-17.

28. Amendt, B.A., E.V. Semina, and W.L. Alward, *Rieger syndrome: a clinical, molecular, and biochemical analysis.* Cell Mol Life Sci, 2000. **57**(11): p. 1652-66.

29. Burian, H.M., M.H. Rice, and L. Allen, *External visibility of the region of Schlemm's canal; report on a family with developmental anomalies of cornea, iris, and chamber angle.* AMA Arch Ophthalmol, 1957. **57**(5): p. 651-8.

30. Henkind, P., I.M. Sigel, and R.E. Carr, *Mesodermal Dysgenesis of the Anterior Segment: Rieger's Anomaly.* Arch Ophthalmol, 1965. **73**: p. 810-7.

31. Sim, K.T., B. Karri, and S.B. Kaye, *Posterior embryotoxon may not be a forme fruste of Axenfeld-Rieger's Syndrome.* J AAPOS, 2004. **8**(5): p. 504-6.

32. Mandal, A.K., V.K. Gothwal, and R. Nutheti, *Surgical outcome of primary developmental glaucoma: a single surgeon's long-term experience from a tertiary eye care centre in India.* Eye, 2007. **21**(6): p. 764-74.

33. Lines, M.A., K. Kozlowski, and M.A. Walter, *Molecular genetics of Axenfeld-Rieger malformations.* Hum Mol Genet, 2002. **11**(10): p. 1177-84.

34. Jorgenson, R.J., et al., *The Rieger syndrome.* Am J Med Genet, 1978. **2**(3): p. 307-18.

35. Feingold, M., et al., *Rieger's syndrome.* Pediatrics, 1969. **44**(4): p. 564-9.

36. Sadeghi-Nejad, A. and B. Senior, *Autosomal dominant transmission of isolated growth hormone deficiency in iris-dental dysplasia (Rieger's syndrome).* J Pediatr, 1974. **85**(5): p. 644-8.

37. Bekir, N.A. and K. Gungor, *Atrial septal defect with interatrial aneurysm and Axenfeld-Rieger syndrome.* Acta Ophthalmol Scand, 2000. **78**(1): p. 101-3.

38. Grosso, S., et al., *Familial Axenfeld-Rieger anomaly, cardiac malformations, and sensorineural hearing loss: a provisionally unique genetic syndrome?* Am J Med Genet, 2002. **111**(2): p. 182-6.

39. De Hauwere, R.C., et al., *Iris dysplasia, orbital hypertelorism, and psychomotor retardation: a dominantly inherited developmental syndrome.* J Pediatr, 1973. **82**(4): p. 679-81.

40. Summitt, R.L., et al., *Mesoectodermal dysplasia of the iris and cornea, mental retardation and myopathy: a sporadic case.* Birth Defects Orig Artic Ser, 1971. **7**(3): p. 129-35.

41. Alward, W.L., *Axenfeld-Rieger syndrome in the age of molecular genetics.* Am J Ophthalmol, 2000. **130**(1): p. 107-15.

42. Walter, M.A., et al., *Autosomal-dominant iridogoniodysgenesis and Axenfeld-Rieger syndrome are genetically distinct.* Ophthalmology, 1996. **103**(11): p. 1907-15.

43. Graff, C., T. Jerndal, and C. Wadelius, *Fine mapping of the gene for autosomal dominant juvenile-onset glaucoma with iridogoniodysgenesis in 6p25-tel.* Hum Genet, 1997. **101**(2): p. 130-4.

44. Mears, A.J., et al., *Autosomal dominant iridogoniodysgenesis anomaly maps to 6p25.* Am J Hum Genet, 1996. **59**(6): p. 1321-7.

45. Kulak, S.C., et al., *Mutation in the RIEG1 gene in patients with iridogoniodysgenesis syndrome.* Hum Mol Genet, 1998. **7**(7): p. 1113-7.

46. Heon, E., et al., *Linkage of autosomal dominant iris hypoplasia to the region of the Rieger syndrome locus (4q25).* Hum Mol Genet, 1995. **4**(8): p. 1435-9.

47. OMIM:, *Online mendelian inheritance in man.OMIM 137600. URL: http://www.ncbi.nlm.nih.gov/omim/.*

48. OMIM:, *Online mendelian inheritance in man, OMIM 180500. URL: http://www.ncbi.nlm.nih.gov/omim/.*

49. Alward, W.L., et al., *Autosomal dominant iris hypoplasia is caused by a mutation in the Rieger syndrome (RIEG/PITX2) gene.* Am J Ophthalmol, 1998. **125**(1): p. 98-100.

50. Brear, D.R. and M.S. Insler, *Axenfeld's syndrome associated with systemic abnormalities.* Ann Ophthalmol, 1985. **17**(5): p. 291-4.

51. Forsius, H., *Visibility of anterior bordering of Schwalbe in congenital anomalities.* Acta Ophthalmology, 1963. **41**: p. 97-107.

52. Alvarez, B., et al., *Forkhead transcription factors contribute to execution of the mitotic programme in mammals.* Nature, 2001. **413**(6857): p. 744-7.

53. Hillion, J., et al., *AF6q21, a novel partner of the MLL gene in t(6;11)(q21;q23), defines a forkhead transcriptional factor subfamily.* Blood, 1997. **90**(9): p. 3714-9.

54. Li, C., et al., *Characterization and chromosomal mapping of the gene encoding the cellular DNA binding protein ILF.* Genomics, 1992. **13**(3): p. 665-71.

55. Medema, R.H., et al., *AFX-like Forkhead transcription factors mediate cell-cycle regulation by Ras and PKB through p27kip1.* Nature, 2000. **404**(6779): p. 782-7.

56. Tran, H., et al., *DNA repair pathway stimulated by the forkhead transcription factor FOXO3a through the Gadd45 protein.* Science, 2002. **296**(5567): p. 530-4.

57. Huang, W., et al., *Nuclear receptor-dependent bile acid signaling is required for normal liver regeneration.* Science, 2006. **312**(5771): p. 233-6.

58. Lin, L., et al., *Modulation of Th1 activation and inflammation by the NF-kappaB repressor Foxj1.* Science, 2004. **303**(5660): p. 1017-20.

59. Nakae, J., et al., *Regulation of insulin action and pancreatic beta-cell function by mutated alleles of the gene encoding forkhead transcription factor Foxo1.* Nat Genet, 2002. **32**(2): p. 245-53.

60. Nemoto, S., M.M. Fergusson, and T. Finkel, *Nutrient availability regulates SIRT1 through a forkhead-dependent pathway.* Science, 2004. **306**(5704): p. 2105-8.

61. Gavin, M.A., et al., *Foxp3-dependent programme of regulatory T-cell differentiation.* Nature, 2007. **445**(7129): p. 771-5.

62. Marson, A., et al., *Foxp3 occupancy and regulation of key target genes during T-cell stimulation.* Nature, 2007. **445**(7130): p. 931-5.

63. Ono, M., et al., *Foxp3 controls regulatory T-cell function by interacting with AML1/Runx1.* Nature, 2007. **446**(7136): p. 685-9.

64. Wan, Y.Y. and R.A. Flavell, *Regulatory T-cell functions are subverted and converted owing to attenuated Foxp3 expression.* Nature, 2007. **445**(7129): p. 766-70.

65. Zheng, Y., et al., *Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells.* Nature, 2007. **445**(7130): p. 936-40.

66. Pierrou, S., et al., *Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending.* EMBO J, 1994. **13**(20): p. 5002-12.

67. Kaestner, K.H., W. Knochel, and D.E. Martinez, *Unified nomenclature for the winged helix/forkhead transcription factors.* Genes Dev, 2000. **14**(2): p. 142-6.

68. Clark, K.L., et al., *Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5.* Nature, 1993. **364**(6436): p. 412-20.

69. Lehmann, O.J., et al., *Fox's in development and disease.* Trends Genet, 2003. **19**(6): p. 339-44.

70. Gould, D.B., et al., *Autosomal dominant Axenfeld-Rieger anomaly maps to 6p25.* Am J Hum Genet, 1997. **61**(3): p. 765-8.

71. Jordan, T., et al., *Familial glaucoma iridogoniodysplasia maps to a 6p25 region implicated in primary congenital glaucoma and iridogoniodysgenesis anomaly.* Am J Hum Genet, 1997. **61**(4): p. 882-8.

72. Larsson, C., et al., *Chromosomal localization of six human forkhead genes, freac-1 (FKHL5), -3 (FKHL7), -4 (FKHL8), -5 (FKHL9), -6 (FKHL10), and -8 (FKHL12).* Genomics, 1995. **30**(3): p. 464-9.

73. Mears, A.J., et al., *Mutations of the forkhead/winged-helix gene, FKHL7, in patients with Axenfeld-Rieger anomaly.* Am J Hum Genet, 1998. **63**(5): p. 1316-28.

74. Nishimura, D.Y., et al., *The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25.* Nat Genet, 1998. **19**(2): p. 140-7.

75. Sheffield, V.C., D.Y. Nishimura, and A.B. Kanis, *Gene dosage of the cause defects of the anterior chamber of the eye.* 1999, Investigative Ophthalmology & Visual Science.

76. Nishimura, D.Y., et al., *A spectrum of FOXC1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye.* Am J Hum Genet, 2001. **68**(2): p. 364-72.

77. Lehmann, O.J., et al., *Chromosomal duplication involving the forkhead transcription factor gene FOXC1 causes iris hypoplasia and glaucoma.* Am J Hum Genet, 2000. **67**(5): p. 1129-35.

78. Lehmann, O.J., et al., *Ocular developmental abnormalities and glaucoma associated with interstitial 6p25 duplications and deletions.* Invest Ophthalmol Vis Sci, 2002. **43**(6): p. 1843-9.

79. Walter, M.A., *PITs and FOXes in ocular genetics: the Cogan lecture.* Invest Ophthalmol Vis Sci, 2003. **44**(4): p. 1402-5.

80. Berry, F.B., et al., *Functional interactions between FOXC1 and PITX2 underlie the sensitivity to FOXC1 gene dose in Axenfeld-Rieger syndrome and anterior segment dysgenesis.* Hum Mol Genet, 2006. **15**(6): p. 905-19.

81. Saleem, R.A., et al., *Analyses of the effects that disease-causing missense mutations have on the structure and function of the winged-helix protein FOXC1.* Am J Hum Genet, 2001. **68**(3): p. 627-41.

82. Berry, F.B., R.A. Saleem, and M.A. Walter, *FOXC1 transcriptional regulation is mediated by N- and C-terminal activation domains and contains a phosphorylated transcriptional inhibitory domain.* J Biol Chem, 2002. **277**(12): p. 10292-7.

83. Berry, F.B., F. Mirzayans, and M.A. Walter, *Regulation of FOXC1 stability and transcriptional activity by an epidermal growth factor-activated mitogen-activated protein kinase signaling cascade.* J Biol Chem, 2006. **281**(15): p. 10098-104.

84. Amendt, B.A., *The Molecular Mechanisms of Axenfeld-Rieger Syndrome.* 2005: Landes Bioscience.

85. Kume, T., et al., *The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus.* Cell, 1998. **93**(6): p. 985-96.

86. Kume, T., K. Deng, and B.L. Hogan, *Murine forkhead/winged helix genes Foxc1 (Mf1) and Foxc2 (Mfh1) are required for the early organogenesis of the kidney and urinary tract.* Development, 2000. **127**(7): p. 1387-95.

87. Kume, T., et al., *The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis.* Genes Dev, 2001. **15**(18): p. 2470-82.

88. Winnier, G.E., et al., *Roles for the winged helix transcription factors MF1 and MFH1 in cardiovascular development revealed by nonallelic noncomplementation of null alleles.* Dev Biol, 1999. **213**(2): p. 418-31.

89. Hong, H.K., J.H. Lass, and A. Chakravarti, *Pleiotropic skeletal and ocular phenotypes of the mouse mutation congenital hydrocephalus (ch/Mf1) arise from a winged helix/forkhead transcriptionfactor gene.* Hum Mol Genet, 1999. **8**(4): p. 625-37.

90. Zarbalis, K., et al., *Cortical dysplasia and skull defects in mice with a Foxc1 allele reveal the role of meningeal differentiation in regulating cortical development.* Proc Natl Acad Sci U S A, 2007. **104**(35): p. 14002-7.

91. Murray, J.C., et al., *Linkage of Rieger syndrome to the region of the epidermal growth factor gene on chromosome 4.* Nat Genet, 1992. **2**(1): p. 46-9.

92. Vaux, C., et al., *Evidence that Rieger syndrome maps to 4q25 or 4q27.* J Med Genet, 1992. **29**(4): p. 256-8.

93. Legius, E., et al., *Genetic heterogeneity in Rieger eye malformation.* J Med Genet, 1994. **31**(4): p. 340-1.

94. Semina, E.V., et al., *Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome.* Nat Genet, 1996. **14**(4): p. 392-9.

95. McGinnis, W., et al., *A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes.* Nature, 1984. **308**(5958): p. 428-33.

96. Scott, M.P. and A.J. Weiner, *Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila.* Proc Natl Acad Sci U S A, 1984. **81**(13): p. 4115-9.

97. Kissinger, C.R., et al., *Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions.* Cell, 1990. **63**(3): p. 579-90.

98. Arakawa, H., et al., *Identification and characterization of the ARP1 gene, a target for the human acute leukemia ALL1 gene.* Proc Natl Acad Sci U S A, 1998. **95**(8): p. 4573-8.

99. Gage, P.J. and S.A. Camper, *Pituitary homeobox 2, a novel member of the bicoid-related family of homeobox genes, is a potential regulator of anterior structure formation.* Hum Mol Genet, 1997. **6**(3): p. 457-64.

100. Kitamura, K., et al., *Mouse Pitx2 deficiency leads to anomalies of the ventral body wall, heart, extra- and periocular mesoderm and right pulmonary isomerism.* Development, 1999. **126**(24): p. 5749-58.

101. Cox, C.J., et al., *Differential regulation of gene expression by PITX2 isoforms.* J Biol Chem, 2002. **277**(28): p. 25001-10.

102. Blum, M., et al., *Vertebrate left-right asymmetry: old studies and new insights.* Cell Mol Biol (Noisy-le-grand), 1999. **45**(5): p. 505-16.

103. Briata, P., et al., *The Wnt/beta-catenin-->Pitx2 pathway controls the turnover of Pitx2 and other unstable mRNAs.* Mol Cell, 2003. **12**(5): p. 1201-11.

104. Chen, C. and M.M. Shen, *Two modes by which Lefty proteins inhibit nodal signaling.* Curr Biol, 2004. **14**(7): p. 618-24.

105. Cheng, S.K., et al., *Lefty blocks a subset of TGFbeta signals by antagonizing EGF-CFC coreceptors.* PLoS Biol, 2004. **2**(2): p. E30.

106. Kioussi, C., et al., *Identification of a Wnt/Dvl/beta-Catenin --> Pitx2 pathway mediating cell-type-specific proliferation during development.* Cell, 2002. **111**(5): p. 673-85.

107. Schweickert, A., et al., *Pitx2 isoforms: involvement of Pitx2c but not Pitx2a or Pitx2b in vertebrate left-right asymmetry.* Mech Dev, 2000. **90**(1): p. 41-51.

108. Shiratori, H., et al., *Two-step regulation of left-right asymmetric expression of Pitx2: initiation by nodal signaling and maintenance by Nkx2.* Mol Cell, 2001. **7**(1): p. 137-49.

109. Tamura, K., S. Yonei-Tamura, and J.C. Belmonte, *Molecular basis of left-right asymmetry.* Dev Growth Differ, 1999. **41**(6): p. 645-56.

110. Weninger, W.J., et al., *Cited2 is required both for heart morphogenesis and establishment of the left-right axis in mouse development.* Development, 2005. **132**(6): p. 1337-48.

111. Amendt, B.A., L.B. Sutherland, and A.F. Russo, *Multifunctional role of the Pitx2 homeodomain protein C-terminal tail.* Mol Cell Biol, 1999. **19**(10): p. 7001-10.

112. Green, P.D., et al., *Antagonistic regulation of Dlx2 expression by PITX2 and Msx2: implications for tooth development.* Gene Expr, 2001. **9**(6): p. 265-81.

113. Hjalt, T.A., et al., *The Pitx2 protein in mouse development.* Dev Dyn, 2000. **218**(1): p. 195-200.

114. Cripps, R.M. and E.N. Olson, *Control of cardiac development by an evolutionarily conserved transcriptional network.* Dev Biol, 2002. **246**(1): p. 14-28.

115. Essner, J.J., et al., *Mesendoderm and left-right brain, heart and gut development are differentially regulated by pitx2 isoforms.* Development, 2000. **127**(5): p. 1081-93.

116. Hjalt, T.A., B.A. Amendt, and J.C. Murray, *PITX2 regulates procollagen lysyl hydroxylase (PLOD) gene expression: implications for the pathology of Rieger syndrome.* J Cell Biol, 2001. **152**(3): p. 545-52.

117. Tremblay, J.J., C.G. Goodyer, and J. Drouin, *Transcriptional properties of Ptx1 and Ptx2 isoforms.* Neuroendocrinology, 2000. **71**(5): p. 277-86.

118. Yu, X., et al., *Differential expression and functional analysis of Pitx2 isoforms in regulation of heart looping in the chick.* Development, 2001. **128**(6): p. 1005-13.

119. Gage, P.J., H. Suh, and S.A. Camper, *Dosage requirement of Pitx2 for development of multiple organs.* Development, 1999. **126**(20): p. 4643-51.

120. Stathacopoulos, R.A., et al., *The Rieger syndrome and a chromosome 13 deletion.* J Pediatr Ophthalmol Strabismus, 1987. **24**(4): p. 198-203.

121. Phillips, J.C., et al., *A second locus for Rieger syndrome maps to chromosome 13q14.* Am J Hum Genet, 1996. **59**(3): p. 613-9.

122. Werner, W., et al., *A small deletion of 16q23.1-->16q24.2 [del(16)(q23.1q24.2).ish del(16)(q23.1q24.2)(D16S395+, D16S348-, P5432+)] in a boy with iris coloboma and minor anomalies.* Am J Med Genet, 1997. **70**(4): p. 371-6.

123. Jamieson, R.V., et al., *Domain disruption and mutation of the bZIP transcription factor, MAF, associated with cataract, ocular anterior segment dysgenesis and coloboma.* Hum Mol Genet, 2002. **11**(1): p. 33-42.

124. Riise, R., K. Storhaug, and K. Brondum-Nielsen, *Rieger syndrome is associated with PAX6 deletion.* Acta Ophthalmol Scand, 2001. **79**(2): p. 201-3.

125. Pauling, L., H.A. Itano, and et al., *Sickle cell anemia a molecular disease.* Science, 1949. **110**(2865): p. 543-8.

126. Pauling, L., H.A. Itano, and et al., *Sickle cell anemia, a molecular disease.* Science, 1949. **109**(2835): p. 443.

127. Ingram, V.M., *A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin.* Nature, 1956. **178**(4537): p. 792-4.

128. Ingram, V.M., *Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin.* Nature, 1957. **180**(4581): p. 326-8.

129. Feuk, L., A.R. Carson, and S.W. Scherer, *Structural variation in the human genome.* Nat Rev Genet, 2006. **7**(2): p. 85-97.

130. Feuk, L., et al., *Structural variants: changing the landscape of chromosomes and design of disease studies.* Hum Mol Genet, 2006. **15 Spec No 1**: p. R57-66.

131. Jacobs, P.A., et al., *The somatic chromosomes in mongolism.* Lancet, 1959. **1**(7075): p. 710.

132. Inoue, K. and J.R. Lupski, *Molecular mechanisms for genomic disorders.* Annu Rev Genomics Hum Genet, 2002. **3**: p. 199-242.

133. McCarroll, S.A. and D.M. Altshuler, *Copy-number variation and association studies of human disease.* Nat Genet, 2007. **39**(7 Suppl): p. S37-42.

134. Lupski, J.R., *Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits.* Trends Genet, 1998. **14**(10): p. 417-22.

135. Lupski, J.R., *Genomic rearrangements and sporadic disease.* Nat Genet, 2007. **39**(7 Suppl): p. S43-7.

136. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome.* Nat Genet, 2004. **36**(9): p. 949-51.

137. Sebat, J., et al., *Large-scale copy number polymorphism in the human genome.* Science, 2004. **305**(5683): p. 525-8.

138. Locke, D.P., et al., *Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome.* Am J Hum Genet, 2006. **79**(2): p. 275-90.

139. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

140. Sharp, A.J., et al., *Segmental duplications and copy-number variation in the human genome.* Am J Hum Genet, 2005. **77**(1): p. 78-88.

141. Wong, K.K., et al., *A comprehensive analysis of common copy-number variations in the human genome.* Am J Hum Genet, 2007. **80**(1): p. 91-104.

142. Stranger, B.E., et al., *Relative impact of nucleotide and copy number variation on gene expression phenotypes.* Science, 2007. **315**(5813): p. 848-53.

143. Lupski, J.R. and P. Stankiewicz, *Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes.* PLoS Genet, 2005. **1**(6): p. e49.

144. Freeman, J.L., et al., *Copy number variation: new insights in genome diversity.* Genome Res, 2006. **16**(8): p. 949-61.

145. Khaja, R., et al., *Genome assembly comparison identifies structural variants in the human genome.* Nat Genet, 2006. **38**(12): p. 1413-8.

146. Sharp, A.J., Z. Cheng, and E.E. Eichler, *Structural variation of the human genome.* Annu Rev Genomics Hum Genet, 2006. **7**: p. 407-42.

147. Tuzun, E., et al., *Fine-scale structural variation of the human genome.* Nat Genet, 2005. **37**(7): p. 727-32.

148. Sachidanandam, R., et al., *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.* Nature, 2001. **409**(6822): p. 928-33.

149. Wang, D.G., et al., *Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome.* Science, 1998. **280**(5366): p. 1077-82.

150. Kaiser, J., *International Team to Sequence 1000 Genomes.* ScienceNOW Daily News 2008.

151. Scherer, S.W., et al., *Challenges and standards in integrating surveys of structural variation.* Nat Genet, 2007. **39**(7 Suppl): p. S7-15.

152. Carter, N.P., *Methods and strategies for analyzing copy number variation using DNA microarrays.* Nat Genet, 2007. **39**(7 Suppl): p. S16-21.

153. Kallioniemi, O.P., et al., *Comparative genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumors.* Semin Cancer Biol, 1993. **4**(1): p. 41-6.

154. Bentley, D.R., et al., *The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X.* Nature, 2001. **409**(6822): p. 942-3.

155. Cheung, V.G., et al., *Integration of cytogenetic landmarks into the draft sequence of the human genome.* Nature, 2001. **409**(6822): p. 953-8.

156. McPherson, J.D., et al., *A physical map of the human genome.* Nature, 2001. **409**(6822): p. 934-41.

157. Solinas-Toldo, S., et al., *Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances.* Genes Chromosomes Cancer, 1997. **20**(4): p. 399-407.

158. Pinkel, D., et al., *High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.* Nat Genet, 1998. **20**(2): p. 207-11.

159. Gribble, S.M., et al., *Ultra-high resolution array painting facilitates breakpoint sequencing.* J Med Genet, 2007. **44**(1): p. 51-8.

160. Gresham, D., M.J. Dunham, and D. Botstein, *Comparing whole genomes using DNA microarrays.* Nat Rev Genet, 2008. **9**(4): p. 291-302.

161. Chen, J.M., et al., *Gene conversion: mechanisms, evolution and human disease.* Nat Rev Genet, 2007. **8**(10): p. 762-75.

162. Lupski, J.R., et al., *DNA duplication associated with Charcot-Marie-Tooth disease type 1A.* Cell, 1991. **66**(2): p. 219-32.

163. Chance, P.F., et al., *DNA deletion associated with hereditary neuropathy with liability to pressure palsies.* Cell, 1993. **72**(1): p. 143-51.

164. Stankiewicz, P. and J.R. Lupski, *Genome architecture, rearrangements and genomic disorders.* Trends Genet, 2002. **18**(2): p. 74-82.

165. Potocki, L., et al., *Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion.* Nat Genet, 2000. **24**(1): p. 84-7.

166. Yobb, T.M., et al., *Microduplication and triplication of 22q11.2: a highly variable syndrome.* Am J Hum Genet, 2005. **76**(5): p. 865-76.

167. Edelmann, L., et al., *A common molecular basis for rearrangement disorders on chromosome 22q11.* Hum Mol Genet, 1999. **8**(7): p. 1157-67.

168. Somerville, M.J., et al., *Severe expressive-language delay related to duplication of the Williams-Beuren locus.* N Engl J Med, 2005. **353**(16): p. 1694-701.

169. Ledbetter, D.H., et al., *Deletions of chromosome 15 as a cause of the Prader-Willi syndrome.* N Engl J Med, 1981. **304**(6): p. 325-9.

170. Kanaar, R., J.H. Hoeijmakers, and D.C. van Gent, *Molecular mechanisms of DNA double strand break repair.* Trends Cell Biol, 1998. **8**(12): p. 483-9.

171. Shaw, C.J. and J.R. Lupski, *Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease.* Hum Mol Genet, 2004. **13 Spec No 1**: p. R57-64.

172. Bailey, J.A., et al., *Segmental duplications: organization and impact within the current human genome project assembly.* Genome Res, 2001. **11**(6): p. 1005-17.

173. Babcock, M., et al., *Shuffling of genes within low-copy repeats on 22q11 (LCR22) by Alu-mediated recombination events during evolution.* Genome Res, 2003. **13**(12): p. 2519-32.

174. Bailey, J.A., et al., *Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22.* Am J Hum Genet, 2002. **70**(1): p. 83-100.

175. Moore, J.K. and J.E. Haber, *Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae.* Mol Cell Biol, 1996. **16**(5): p. 2164-73.

176. Walker, J.R., R.A. Corpina, and J. Goldberg, *Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair.* Nature, 2001. **412**(6847): p. 607-14.

177. Roth, D.B., T.N. Porter, and J.H. Wilson, *Mechanisms of nonhomologous recombination in mammalian cells.* Mol Cell Biol, 1985. **5**(10): p. 2599-607.

178. Roth, D.B. and J.H. Wilson, *Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction.* Mol Cell Biol, 1986. **6**(12): p. 4295-304.

179. Raghavan, S.C., J. Tong, and M.R. Lieber, *Hybrid joint formation in human V(D)J recombination requires nonhomologous DNA end joining.* DNA Repair (Amst), 2006. **5**(2): p. 278-85.

180. Schwartz, M., et al., *Homologous recombination and nonhomologous end-joining repair pathways regulate fragile site stability.* Genes Dev, 2005. **19**(22): p. 2715-26.

181. Sonoda, E., et al., *Differential usage of non-homologous end-joining and homologous recombination in double strand break repair.* DNA Repair (Amst), 2006. **5**(9-10): p. 1021-9.

182. Lieber, M.R., et al., *Mechanism and regulation of human non-homologous DNA end-joining.* Nat Rev Mol Cell Biol, 2003. **4**(9): p. 712-20.

183. Tsukamoto, Y. and H. Ikeda, *Double-strand break repair mediated by DNA end-joining.* Genes Cells, 1998. **3**(3): p. 135-44.

184. Nobile, C., et al., *Analysis of 22 deletion breakpoints in dystrophin intron 49.* Hum Genet, 2002. **110**(5): p. 418-21.

185. Inoue, K., et al., *Genomic rearrangements resulting in PLP1 deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females.* Am J Hum Genet, 2002. **71**(4): p. 838-53.

186. Woodward, K.J., et al., *Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination.* Am J Hum Genet, 2005. **77**(6): p. 966-87.

187. Heilstedt, H.A., et al., *Physical map of 1p36, placement of breakpoints in monosomy 1p36, and clinical characterization of the syndrome..* Am J Hum Genet, 2003. **72**(5): p. 1200-12.

188. de Mollerat, X.J., et al., *A genomic rearrangement resulting in a tandem duplication is associated with split hand-split foot malformation 3 (SHFM3) at 10q24.* Hum Mol Genet, 2003. **12**(16): p. 1959-71.

189. Joyce, G.F., *Molecular evolution: booting up life.* Nature, 2002. **420**(6913): p. 278-9.

190. Fouche, N., et al., *Replication fork regression in repetitive DNAs.* Nucleic Acids Res, 2006. **34**(20): p. 6044-50.

191. Mirkin, E.V. and S.M. Mirkin, *Replication fork stalling at natural impediments.* Microbiol Mol Biol Rev, 2007. **71**(1): p. 13-35.

192. Lambert, S., et al., *Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier.* Cell, 2005. **121**(5): p. 689-702.

193. Lee, J.A., C.M. Carvalho, and J.R. Lupski, *A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders.* Cell, 2007. **131**(7): p. 1235-47.

194. Burd, L., et al., *Partial 6p trisomy associated with infantile autism.* Clin Genet, 1988. **33**(5): p. 356-9.

195. Davies, A.F., et al., *A detailed investigation of two cases exhibiting characteristics of the 6p deletion syndrome.* Hum Genet, 1996. **98**(4): p. 454-9.

196. Song, M. and L. Li, *Partial trisomy 6p due to maternal t(1;6) translocation.* Clin Genet, 1996. **49**(6): p. 316-7.

197. Davies, A.F., et al., *Evidence of a locus for orofacial clefting on human chromosome 6p24 and STS content map of the region.* Hum Mol Genet, 1995. **4**(1): p. 121-8.

198. Gould, D.B., et al., *Phenotypic and molecular assessment of seven patients with 6p25 deletion syndrome: relevance to ocular dysgenesis and hearing impairment.* BMC Med Genet, 2004. **5**: p. 17.

199. Le Caignec, C., et al., *Subtelomeric 6p deletion: clinical, FISH, and array CGH characterization of two cases.* Am J Med Genet A, 2005. **132**(2): p. 175-80.

200. Maclean, K., et al., *Axenfeld-Rieger malformation and distinctive facial features: Clues to a recognizable 6p25 microdeletion syndrome.* Am J Med Genet A, 2005. **132**(4): p. 381-5.

201. Ekong, R., et al., *Chromosomal anomalies on 6p25 in iris hypoplasia and Axenfeld-Rieger syndrome patients defined on a purpose-built genomic microarray.* Hum Mutat, 2004. **24**(1): p. 76-85.

202. Descipio, C., et al., *Subtelomeric deletions of chromosome 6p: molecular and cytogenetic characterization of three new cases with phenotypic overlap with Ritscher-Schinzel (3C) syndrome.* Am J Med Genet A, 2005. **134**(1): p. 3-11.

203. Koolen, D.A., et al., *Partial iris hypoplasia in a patient with an interstitial subtelomeric 6p deletion not including the forkhead transcription factor gene FOXC1.* Eur J Hum Genet, 2005. **13**(11): p. 1169-71.

204. Hong, H.K., et al., *The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice.* Genesis, 2001. **29**(4): p. 163-71.

205. Aitola, M., et al., *Forkhead transcription factor FoxF2 is expressed in mesodermal tissues involved in epithelio-mesenchymal interactions.* Dev Dyn, 2000. **218**(1): p. 136-49.

Chapter 2:

A novel mechanistic spectrum underlies glaucoma associated chromosome 6p25 copy number variations

A version of this chapter has been submitted to the

American Journal of Human Genetics.

Bhaskar Chanda, Mika Asai-Coakwell, Ming Ye, Andrew J Mungall, Margaret Barrow, Jane C. Sowden, Nigel P. Carter, Michael A.Walter and Ordan J. Lehmann

Summary

The factors that mediate chromosomal rearrangement remain incompletely defined. Amongst regions prone to structural variant formation, chromosome 6p25 is one of the few in which disease-associated segmental duplications and segmental deletions have been identified, primarily through gene dosage attributable ocular phenotypes. We studied the largest collection of 6p25 duplication and deletion pedigrees worldwide, and amplified junctional fragments from each. Analysis of the breakpoint architecture revealed that in contrast to previous examples all the rearrangements were non-recurrent, and that most utilized coupled homologous and non-homologous recombination mechanisms. One junctional fragment, with an unprecedented 367 base pair insert derived from tandemly arranged breakpoint elements, accorded with a recently described replication based mechanism. However, it differed from the sole previous example in not being associated with template switching, and occurring in a segmental deletion. These results extend the mechanisms involved in structural variant formation, provide strong evidence that a spectrum of recombination, DNA repair and replication underlie 6p25 rearrangements, and have implications for genesis of copy number variations in other genomic regions. These findings highlight the benefits of undertaking the extensive studies necessary to characterize CNVs at the base pair level.

Introduction

The contribution that chromosomal anomalies make to human disease was, until relatively recently, under-appreciated. Introduction of array technologies has revealed that rearrangements below the resolution of standard karyotyping are exceedingly common [1, 2] with sub-microscopic structural genomic variants underlying a substantial proportion of human genetic variation [3, 4] and disease [5-8]. Estimates that such variants account for 12% or more of our genomes [2, 9] are likely conservative due to the resolution of the arrays used, and these technologies' general inability to detect most balanced rearrangements [9]. Despite their importance to a wide range of human diseases [5-8], structural variants have been defined at the base pair level in only a small number of cases, so restricting our understanding of the underlying mechanisms involved. As the majority of characterized rearrangements were identified prior to the advent of comparative genomic hybridization (CGH) and single nucleotide polymorphism (SNP) arrays, it remains to be determined whether the mechanisms involved in their formation are representative of the generally smaller genomic variants being detected in large numbers with array approaches.

Since copy number alteration predominates amongst the mechanisms by which genomic rearrangements induce phenotypes [10], copy number variation (CNV) is

becoming the preferred descriptor for segmental gains or losses of chromosomal material. CNVs have particular relevance in ocular genetics due to the eye's sensitivity to the effects of altered gene dosage [11]. Combined with the organ's accessibility to detailed phenotyping, these factors provide unique opportunities to use CNV-induced phenotypes to elucidate the functions of dosage sensitive genes [12-15]. Amongst dosage-sensitive ocular developmental genes, the forkhead box transcription factor *FOXC1* (MIM 601090), located on chromosome 6p25, is noteworthy for the frequency of its involvement in segmental and telomeric chromosomal rearrangements [12, 13, 16-19]. Increased or decreased *FOXC1* copy number causes the Axenfeld-Rieger spectrum of glaucoma-associated ocular anomalies [ARS, (MIM 109120)] [12, 13, 17, 19], and occurs at a similar prevalence to *FOXC1* mutation [20]. Although previous studies localized 6p25 segmental anomalies using microsatellite marker genotyping or large-insert bacterial clones as probes for FISH, the resulting broad intervals (150-200kb) precluded determining the causation of these rearrangements [17, 21, 22]. Further investigation was warranted by the unique architecture of 6p25, which contains a triplet of tandemly arranged paralogs (*FOXC1*, *FOXF2*, *FOXQ1*) (MIM 602402, 601705), and thus differs from the five other regions where segmental duplications and deletions cause human disease {Charcot-Marie-Tooth Disease [23] [dup17(p11.2-12)] and Hereditary Neuropathy with liability to pressure palsies [24] [del17(p11.2)]; Potocki-Lupski [dup17(p11.2p11.2)] and Smith Magenis syndromes [25] [del17(p11.2p11.2)]; dup22(q11.2) and velocardiofacial syndrome [26, 27];

dup7(q11.23) and Williams-Beuren Syndrome [28]; dup15(q11-13) [29] and Prader-Willi/Angelman Syndromes [30]}.

Accordingly, in order to characterize the 6p25 rearrangements at the base pair level, amplicon-based, and subsequently commercial oligonucleotide CGH arrays, were used to comprehensively analyze the 6p25 region in the largest ARS pedigree collection worldwide [13, 17, 19-21]. The findings revealed a novel mechanistic spectrum involved in the genesis of these CNVs, including junction fragment architecture incompatible with the two primary mechanisms [non-allelic homologous recombination (NAHR) and non-homologous end joining (NHEJ)] that underlie characterized genomic rearrangements. These results broaden the mechanisms that generate copy number variation with implications for CNV formation in other parts of the genome.

Materials and Methods

Patients

Genomic DNA from 148 individuals with ARS phenotypes secondary to 6p25 rearrangements (Table 2.1), were analyzed after isolation from venous blood samples using conventional techniques [31, 32]. These samples were derived from seven segmental duplication [13, 33], and one segmental deletion pedigree [17], as well as a single ring chromosome proband. The ancestral UK origin of the duplication pedigrees, recruited from Canada (n=3) and the UK (n=4), was compatible with the existence of founder effects (Table 2.1). A panel of 32 mutation-negative ARS cases, previously screened by microsatellite marker genotyping for copy number changes involving the known ARS-associated genes *FOXC1* and *PITX2* (MIM 601542), were also studied with *FOXC1* quantitative PCR [TaqMan, ABI]. The study was approved by the Health Research Ethics Board of the University of Alberta.

Analysis of 6p25 Copy Number Variants

DNA from 4 segmental duplication patients were initially hybridized to CGH arrays comprising 544 contiguous 744 – 10289 bp PCR amplicon-based probes from a 680 kb region encompassing the segmental duplications. Amplicon design and

Table 2.1. Ocular phenotypes of 6p25 copy number variation families

Pedigree	Phenotype	Location and Reference(s)
#1	Iris hypoplasia and glaucoma	Canada [33-35]
#2	Iris hypoplasia and glaucoma	Canada
#3	Iris hypoplasia and glaucoma	UK [22]
#4	Iris hypoplasia and glaucoma	UK
#5	Iris hypoplasia and glaucoma	UK [13]
#6	Iris hypoplasia and glaucoma	Canada [33, 34]
#7	Iris hypoplasia and glaucoma	UK [22, 35]
#8	Axenfeld-Reiger Syndrome and glaucoma	UK [17, 22]
#9	Axenfeld-Rieger Syndrome	Canada
#10	Anterior segment dysgenesis	UK

[Note. The ancestral origin of #1, 2 and 6 is from the UK]

hybridizations were performed as previously described [36]. Subsequently, a customized commercial array comprising isothermal, oligonucleotide probes tiled at a high (~100 bp) density, across the telomeric 5 Mb of 6p25 was employed using Cy3-labeled patient and Cy5-labeled reference DNA as described elsewhere [37]. DNA samples from a single affected individual from seven pedigrees [#1 [35], 3 [22], 4, 5 [13] and 7 [17, 35] (duplication), #8 (deletion) [17, 22], #10 (ring chromosome 6)] and an ARS case (#9) in which quantitative PCR (qPCR) was compatible with a CNV, were hybridized using standard techniques to determine the extent of the 6p25 rearrangements. After an *Alu*-based PCR approach [38] (n= 20 primer pairs) proved unsuccessful at amplifying breakpoint-spanning junction fragments, additional primer pairs (n=110) extending 20 kb from the predicted breakpoints were designed. Since array CGH cannot define the relative positions or orientation of the duplicated and unduplicated segments, long–range PCR (Elongase, Invitrogen) was next performed with multiple primer permutations (n=336) to generate junction fragments in all the segmental duplication pedigrees (Table 2.2). In one pedigree (#7), supplementary primers (n=48) extending up to 70 kb from the predicted breakpoints, were required. Junction fragments were visualized on ethidium bromide stained 1% agarose gels and sequenced with internal primers (Table 2.2) using a BigDye (v3.1) terminator kit and 3100 DNA sequencer (Applied Biosystems). In pedigree #8, the amplicon's architecture necessitated cloning into a vector (pCR®4-TOPO®, Invitrogen) prior to sequencing.

Bioinformatic and haplotype analysis

In silico analyses of sequences adjacent to the breakpoints were performed with Ensembl [University of California–Santa Cruz (UCSC)] and National Center for Biotechnology Information (NCBI) Entrez genome browsers. Sequence composition was determined with BLASTN (NCBI Blast) and BLAT multi-species comparison (UCSC), with the boundaries of repeat elements defined using PipMaker, Ensembl and UCSC. To determine whether the duplication pedigrees shared common haplotypes, genotyping was performed with known (*D6S967, RH122719, SHGC-82115, SHGC-53095* and *SHGC-112337*) or predicted fluorescent labeled microsatellite markers (*BA121, BA118* and *FF2*) identified by analysis of repeat elements in the duplicated segments, supplemented by analysis of 11 SNPs (rs12524120/4544/5352/5740; rs1961687; rs1785778/5836/6028/6106 together with 2 SNPs identified within an E74-like factor pseudogene, SHGC-149664).

Table 2.2. Primers (5'→3') for amplifying junctional fragments, *FOXC1* qPCR and haplotype analysis.

	Forward	Reverse	Annealing Temp (°C)
Junctional Fragment			
#1 – 2	GAACTTGTAACATTCCCAACAGTTCT	CATGTGTGACAAGTGATATAAGCAAG	60
#3 – 6	TACAGAACAGACAGTACAGATTATGAGG	GGAGATGGGTCATAAACATTTGAACTATAT	60
#7	TGCTTAGGTCTATAGGACAGAGTCG	CCAAGAGAGTCTCCTTTCATTGTTA	60
#8	TCAAACATTGAGGGTAGTGTTTTGG	CCATGTATTGTGCACACACA	60
#9	AGAGCCTCTCTTCTGTTTAAGACATC	CCAGAAACTCTTCCTAAGAGAAAGAA	62
qPCR			
#9	CACTCGGTGCGGGAGATG Probe: TCGAGTCACAGAGGAT	TGAACAACTCTCCAGTGAACGG	60
Microsatellite markers/SNPs			
BA121	CAGGAGGCAGAGGTCACAGT	ATAGGGTTGGTTCCCAAAGG	60
FF2	CGGGTTCCACCTACATGGC	CACTCCAGCATGTCCTCCTACT	58
E74 SNP1+2	GTGGCTAGATTTGTGAGTATCACTTC	CGAGTTATCGGTGGTAGTCATACA	58

Results

We hybridized DNA samples from affected individuals in the majority of our pedigrees to CGH arrays. Initially an in-house PCR amplicon-based array was used, which provided clear hybridization data. However, the limited extent of this array, not encompassing both breakpoints in pedigrees #7 and #10 (data not shown), was a factor in subsequent adoption of a commercial array. The latter accurately defined the extent of the rearrangements in 7 of the 8 pedigrees [duplication: #1, 3-5; deletion: #8 and 9; ring: #10] in which it was used (Figure 2.1). Alu-based PCR, adopted initially failed to amplify any junction fragment. Subsequently, long-range PCR was adopted to amplify junction fragments in six of the seven segmental duplication pedigrees (#1-6), as well as both deletion pedigrees (#8 and 9). The identical junction fragment sizes observed in the duplication pedigrees #1-2 (2.5 kb) and #3-6 (2.1 kb), were compatible with either recurrent rearrangements or founder effects.

Sequence analysis of the junction fragment from pedigrees #1 and 2 revealed identical head to tail arranged 492 kb duplicons, separated by a 6 bp insertion (Figure 2.2A). The telomeric breakpoint lies within a long interspersed nuclear element (LINE), located within 3 kb of near-contiguous repeats, whilst the centromeric breakpoint lies 381 bp away from a low complexity region (LCR) that is located within a 10 kb region of repetitive sequence. The maximum repeat

Figure 2.1 Characterization of 6p25 CNVs

(A-B) Montage of array CGH results from patients with Axenfeld-Rieger phenotypes illustrating the extent of the structural genomic variants relative to the position of the *forkhead* paralogs and other genes. Duplicated or deleted genes are depicted in brown (others in grey): note the constant (*FOXC1*) and varying (*FOXF2 / FOXQ1*) involvement in the rearrangements.

A. The segmental duplications extend 492 (*A*,I pedigrees #1-2), 480 (*A*,II #3-6) and 529kb respectively (*A*,III #7), the segmental deletions (*A*,IV #8 and *A*,V #9) 1216 and 30kb.

B. The ring chromosome encompasses 2452 kb of copy number change on 6p25 & 6q27.

C. Schematic overview of the extent of each 6p25 chromosomal anomaly (red and open arrows respectively).

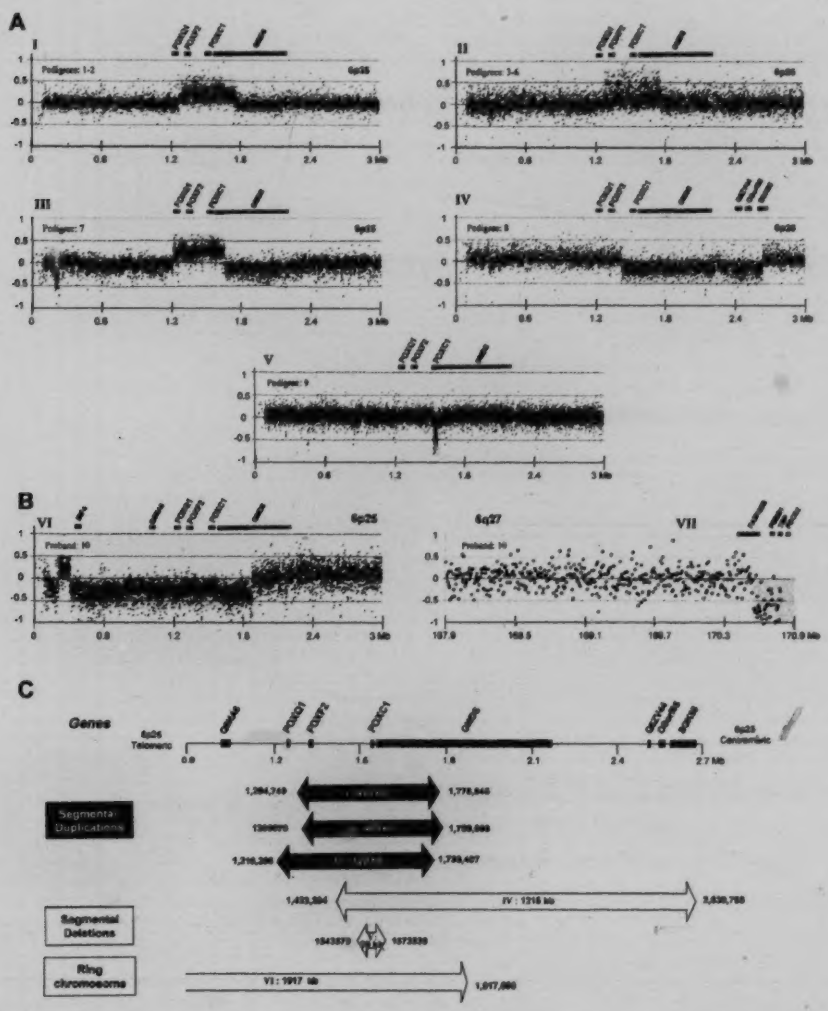
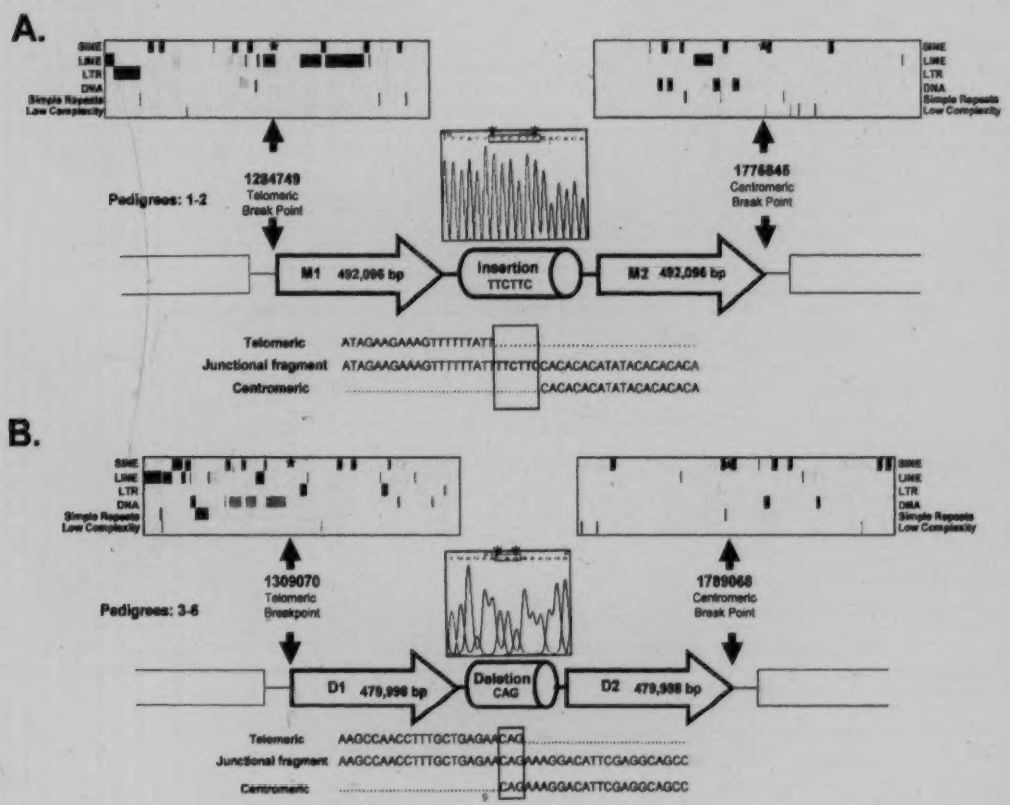


Figure 2.2 Summary of genomic architecture in 6p25 segmental duplication pedigrees (#1-6)

The genomic architecture of the 20 kb region flanking the breakpoints is shown above the respective breakpoints, with the distribution of repeat elements inset above (UCSC Genome Bioinformatics).

A. Pedigrees (#1-2) have an identical head to tail arrangement of a 492,096 bp duplicated segment separated by a 6bp insert. The telomeric breakpoint (red star) lies within a LINE while the centromeric breakpoint (blue star) lies 381 bp from a low copy repeat.

B. Pedigrees (#3-6) share similarly oriented 479,998 bp duplicated motifs separated by a 3 bp deletion. The telomeric breakpoint lies 191 bp from a MER1_type repeat while the centromeric breakpoint lies 381bp from an AT-rich low complexity repeat.

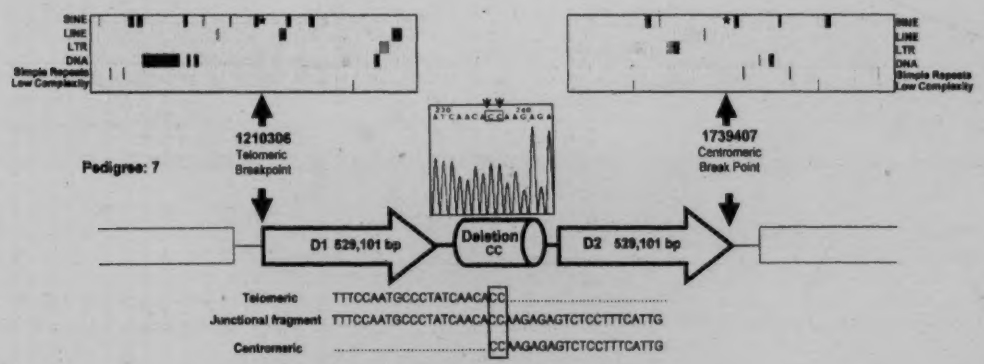


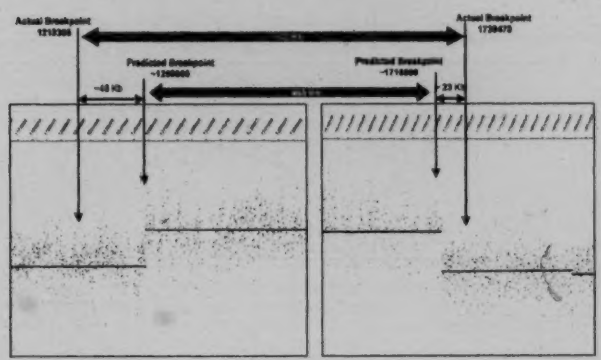
element composition in a 10 kb region encompassing the telomeric and centromeric breakpoints is 71% and 45% respectively, compared to ~35% genome-wide [39]. Four pedigrees (#3-6) share similarly oriented, but smaller 480 kb duplicated segments, separated by a 3 bp deletion (Figure 2.2B). The breakpoints, located ~ 25 kb from those in #1 and 2, lie close to repeat elements [telomeric: MER1_type repeat (191 bp); centromeric: between two short interspersed nuclear elements (SINE), AluSx and MIRb respectively]. The 10 kb region adjacent to the telomeric breakpoint is noteworthy for the high (94%) repeat density (compared to centromeric 28%), indeed a near contiguous (97%) 13.8 kb repeat block lies adjacent to the telomeric breakpoint.

Despite clear array CGH based prediction of the breakpoint positions in the 7th duplication pedigree (Figure 2.3), no junction fragment was initially amplified. To assess whether the duplicons were separated by more than 20 kb, additional primers extending up to 70 kb from the predicted breakpoints were used. One primer combination yielded a junction fragment, with subsequent sequencing revealing significant discrepancies (Figure 2.3) between the actual and predicted breakpoints [telomeric 46 kb and centromeric 23 kb]. Whilst these intervals had no significant homology to other chromosomal regions, the telomeric interval exhibited a 40% repetitive sequence composition (centromeric 15%). Pedigree #7's head to tail orientated 529 kb duplicated segment (Figure 2.3) is larger than the previous examples and exhibits a 2 bp deletion at the junction and the breakpoints lie adjacent

Figure 2.3 Montage illustrating the genomic architecture of the 7th segmental duplication.

The actual breakpoints, and those predicted by array CGH, shown below. The head to tail orientated 529,101bp duplicated segment is separated by 2 bp deletion, with the breakpoints lying adjacent to *Alu* elements. Note the discrepancies in the positions of the actual and predicted breakpoints, in the presence of very clear CGH data.





to AluJo and AluJb repeat elements (telomeric 47 bp, centromeric 735 bp) that share 90% (135/150) sequence homology. The prevalence of repeat elements in the telomeric and centromeric 10 kb regions encompassing the breakpoints is 59% and 42% respectively.

In contrast to the 480-529 kb duplications (#1-7), segmental deletion #8 is 1.22 Mb in size with breakpoints separated by insertion of an unexpectedly large 367 bp of novel sequence. This insert comprises two 100% homologous motifs (M1 & M2), separated by a 13 bp DNA segment (Figure 2.4A). Motif M1 consists of 3 portions, the largest of which at 128 bp of primarily (GTG)n repeats is revealed by BLAT analysis to be homologous to three adjacent segments of sequence spanning the centromeric breakpoint (Figure 2.4B: I, II, III). The remaining motif portions of 15 and 83 bp, are wild type sequence that respectively lie at the telomeric and centromeric breakpoints (Figure 2.4A). Motif M2 comprises 211 bp that is identical to M1 together with the same 15 bp of telomeric wild type sequence (grey, Figure 2. 4A,C). The breakpoints are located within similarly sized (GTG)n simple repeats (telomeric 183 bp, centromeric 180 bp) with the former adjacent to three (GTG)n or (ATGGTG)n elements that extend for a near contiguous 889 bp. In addition, topoisomerase I consensus cleavage (CAT/GTC), and DNA polymerase α pause sites (GAG) are present at or adjacent to both regions (Figure 2.4B).

Quantitative PCR screening of an ARS patient panel unattributable to *FOXC1* or *PITX2* mutation, identified one sample with reduced *FOXC1* copy number. This was

Figure 2.4 Summary of the segmental deletion's (#8) genomic architecture.
A. Summary of the segmental deletion's (#8) genomic architecture illustrating the composition of the 226bp duplicated motifs (M1, M2), with the 20 kb region flanking each breakpoint (filled arrows) shown above. The motifs comprise novel (green) and wild type sequence (grey), with the pattern of shading used to highlight the 100% homology of these motifs.
B. Illustration of the 100% sequence homology between motif segments and regions adjacent to both breakpoints: the location of DNA polymerase α pause sites and consensus cleavage sites for Topo-isomerase I are depicted with red and blue stars respectively. Note the corresponding composition, order and orientation of the 128bp motif portion and segments of sequence (I, II, III) that straddle the centromeric breakpoint (identical bases in capitals).
C. Illustration of the motifs' 100% sequence homology (text colour corresponding to *A*), 5bp sequence at end of novel M2 motif sequence underlined. Below, diagramatic representation of motifs illustrating the frequent interspersed trinucleotide (GTG) repeats (filled circle) separated by multiples of n=3 nucleotides, except at the 3'end of unique sequence in motif 2 (underlined).

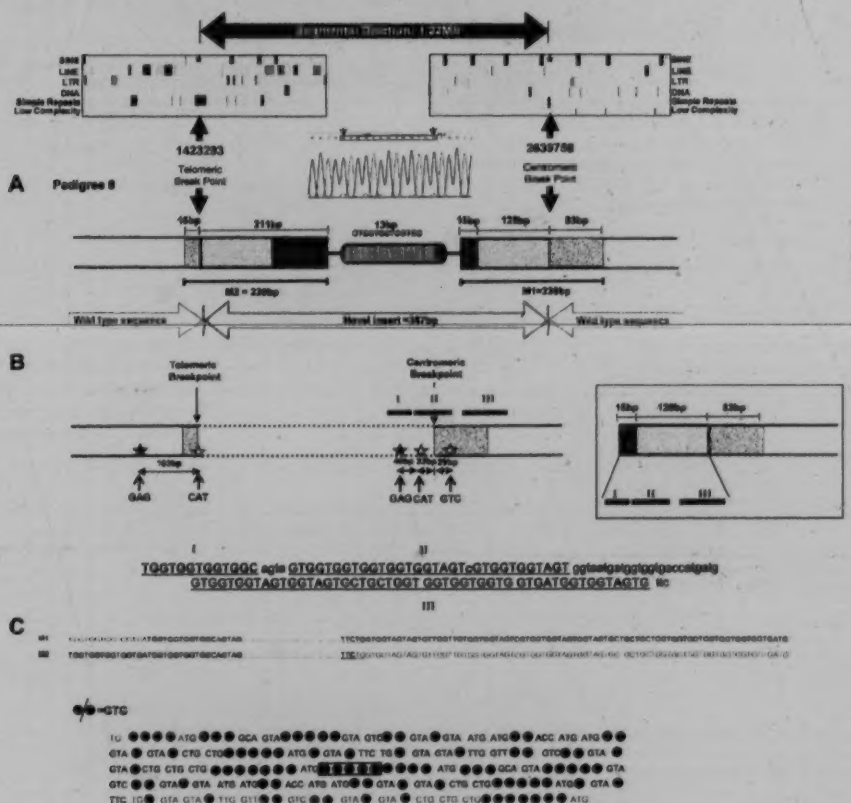
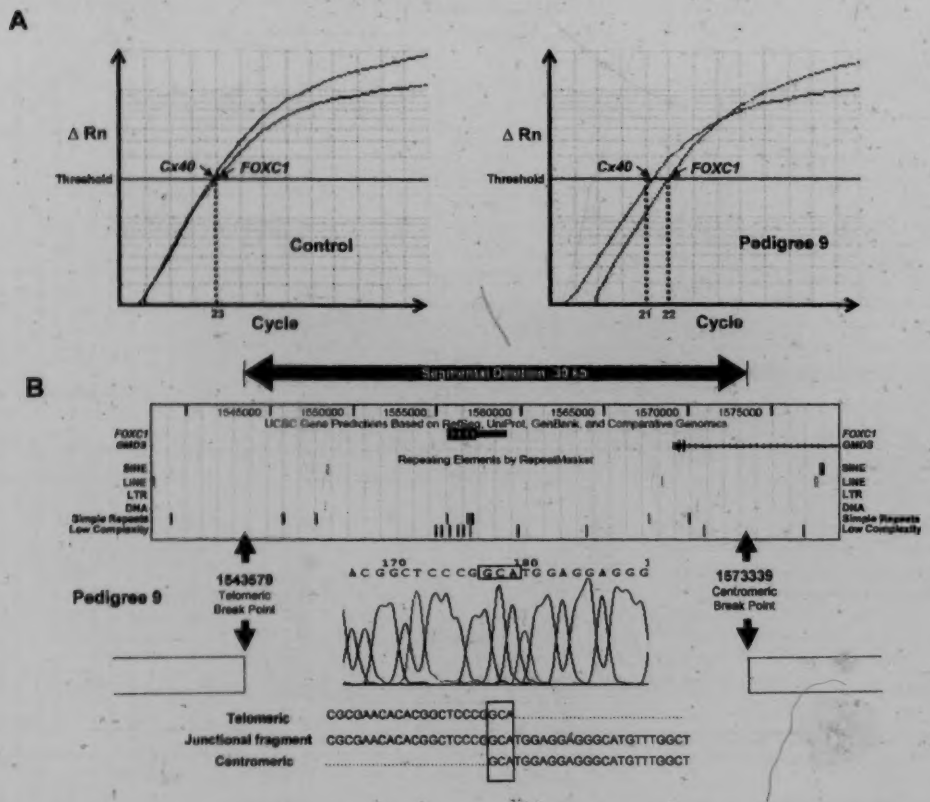


Figure 2.5 Schematic representation of qPCR-identified segmental deletion (#9).

A. Note the lag in the qPCR's exponential phase, with *FOXC1* reaching threshold one cycle after the *Connexin40* (*Cx40*) control. The exponential phase (threshold) of the qPCR is denoted by the red line. Solid black arrows indicate graphical lines corresponding to *FOXC1* and *Connexin40*.
B. The architecture of the segmental deletion's breakpoint is shown below.



confirmed with array CGH, defining a small segmental deletion (29.76 kb) (Figure 2.5) in contrast to the previous much larger duplication CNVs (~480-1220 kb). Sequencing of the junction fragment demonstrated a deletion of 3 bp, with the 10kb regions around the breakpoints noted to be remarkably devoid of repeat elements (2.2% and 0% respectively) (Figure 2.5). This deletion encompasses *FOXC1* together with a small part of GDP-mannose 4,6-dehydratase (*GMDS,* MIM 602884).

The junction fragments segregated with the disease phenotype in all the segmental anomaly pedigrees (Figure 2.6 *A-C* and unpublished data), and the absence of a junction fragment in either parent of the proband (#9) is compatible with a *de-novo* origin in this pedigree (Figure 2.6D). Five of 19 microsatellite markers or SNPs studied were informative (Table 2.3), demonstrating that pedigrees #1 and 2, and separately #3 - 6 share common haplotypes across the duplicated segment, in addition to identically sized segmental duplications and breakpoint architectures [#1 and 2: 492,096 bp with a 6 bp insertion; #3 - 6: 479,998 bp with a 3 bp deletion] (Figures 2.7A-B). Array CGH of #10 revealed loss of chromosomal material from the telomeres of 6p and 6q with three regions of copy number variation identified (Figure 2.1B). The inability to amplify a junction fragment may be attributable to this rearrangement's complexity. Although differing numbers of forkhead genes are encompassed by each rearrangement [duplication (*FOXC1* and *FOXF2*: 7/7; *FOXQ1*: 1/7), deletion (*FOXC1*: 2/2), ring (*FOXC1*, *FOXF2*, *FOXQ1*)] all 7 duplication pedigrees exhibit similar ocular structural phenotypes (Table 2.1).

Figure 2.6 Montage illustrating the segregation of junctional fragments with disease phenotype in representative pedigrees.

A,B. Segmental duplications (#1 and #4, respectively), *C, D.* Segmental deletion (#8 and #9, respectively); note the unaffected parents of individual 2 (probands denoted by arrowhead).

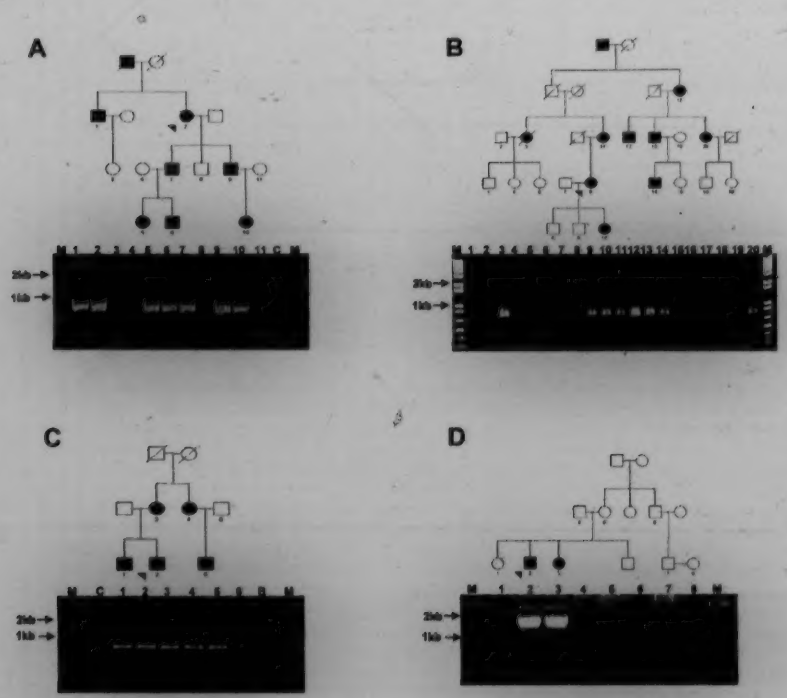


Figure 2.7.A Microsatellite marker genotyping in pedigree #1.
The alleles, corresponding to the genomic region encompassed by segmental duplication in pedigree #1, are indicated with open and hashed bars. Note the hashed bar, representing the duplicated allele, segregates with the disease phenotype. On the top right hand corner, the relative positions of the microsatellites markers (*D6S967*, *BA121* and *FF2*), SNPs (SNP1 and SNP2) and breakpoints of the segmental duplication are depicted. The proband is indicated with a solid black arrow.

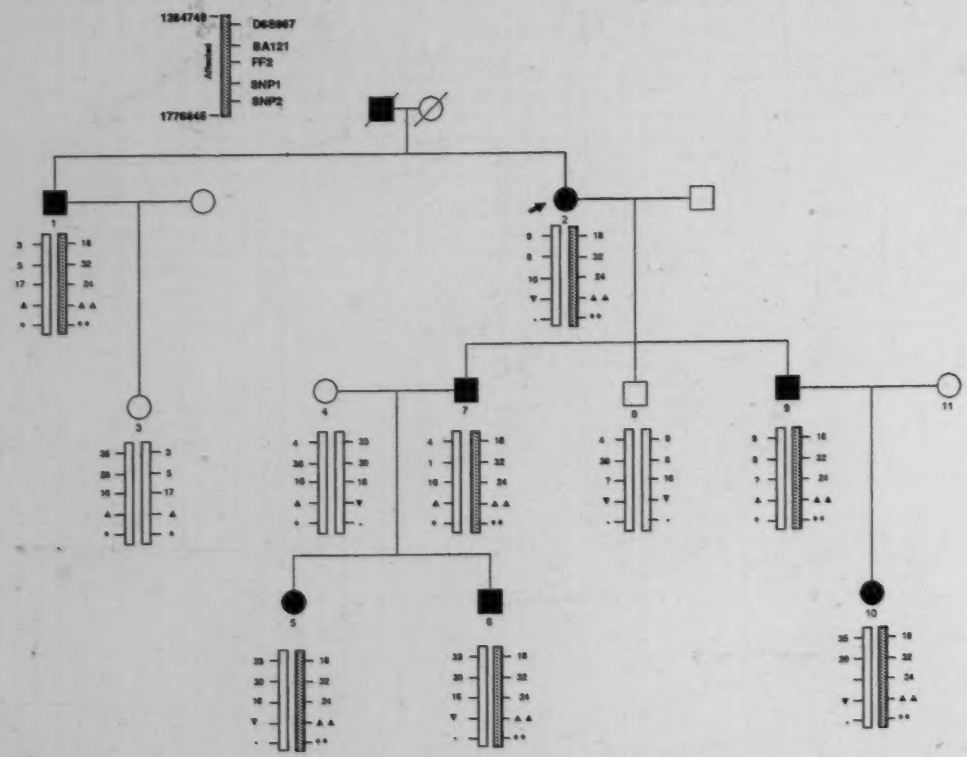


Figure 2.7.B Microsatellite marker genotyping in pedigree #4.
The alleles, corresponding to the genomic region encompassed by segmental duplication in pedigree #4, are indicated with open and hashed bars. Note the hashed bar, representing the duplicated allele, segregates with the disease phenotype. On the top right hand corner, the relative positions of the microsatellites markers (*D6S967, BA121* and *FF2*), SNPs (SNP1 and SNP2) and breakpoints of the segmental duplication are depicted. The proband is indicated with a solid black arrow.

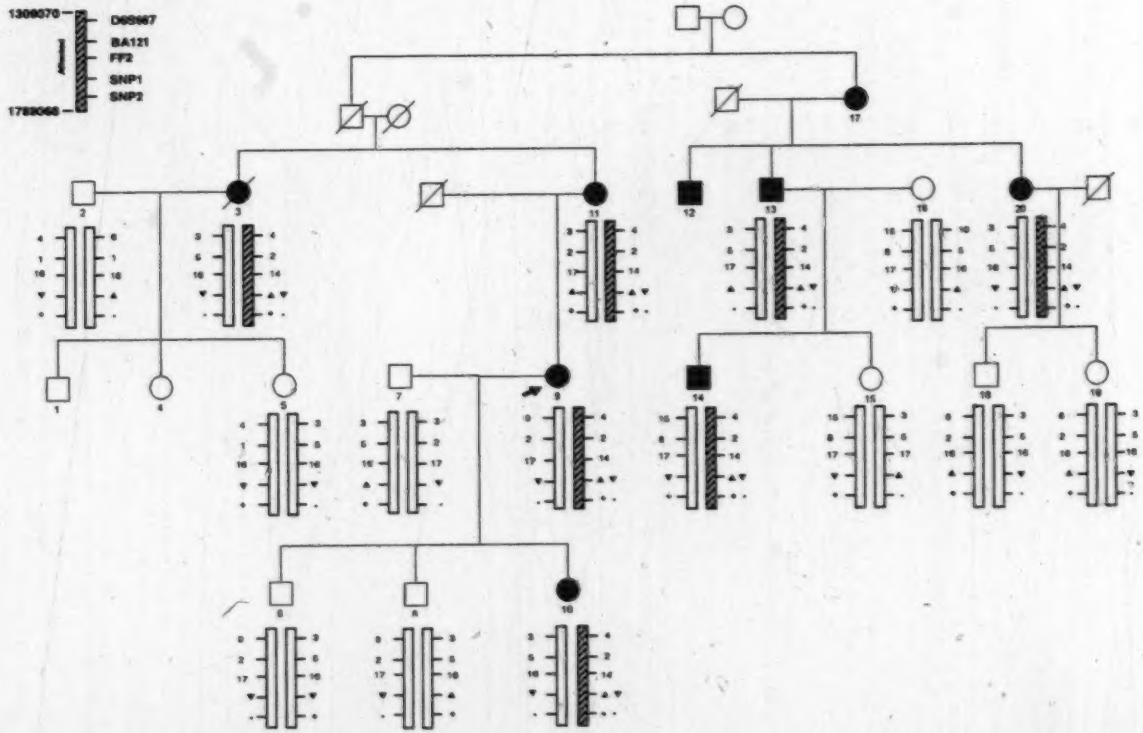


Table 2.3. Haplotype of segmental duplication pedigrees

Pedigrees	Microsatellite Markers			SNPs		Architecture	
	D6S967	*BA121*	*FF2*	*SNP-1*	*SNP-2*	Extent	Insertion/Deletion
1	18	32	24	▲ ▲	♦ ♦	492,096 bp	6bp insertion
2	18	32	24	▲ ▲	♦ ♦	492,096 bp	6bp insertion
3	4	2	14	▲ ▼	♦ *	479,998 bp	3bp deletion
4	4	2	14	▲ ▼	♦ *	479,998 bp	3bp deletion
5	4	2	14	▲ ▼	♦ *	479,998 bp	3bp deletion
6	4	2	14	▲ ▼	♦ *	479,998 bp	3bp deletion

Discussion

Elucidating the genesis of CNVs is important in view of the major contributions that they make to human disease and the fact that despite recent advances [40], the mechanisms that mediate their formation remain incompletely defined. The recurrent nature of most characterized disease-associated segmental rearrangements illustrates that they arise due to the neighbouring genomic architecture. Such (recurrent) structural variants, which have common breakpoints in unrelated individuals, are usually generated by low-copy repeats that induce misalignment during meiosis and thus provide the substrate for NAHR [41-44]. Less frequently, rearrangements occur in which the breakpoint positions vary, and these non-recurrent rearrangements are generally attributable to NHEJ – a process characterized by insertion or deletion of nucleotides between breakpoints lying in non-homologous regions [45-50]. One novel aspect of the present study, stems from the broader mechanistic spectrum observed, as evident from the following features.

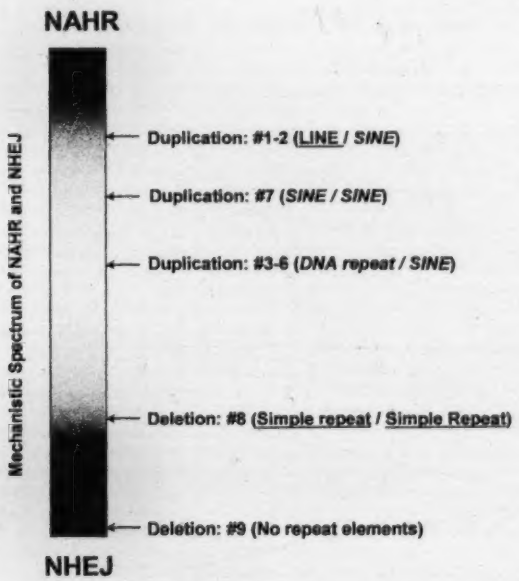
Firstly, previous examples of disease-associated duplication and deletion in the same genomic area predominantly have common breakpoints [51-53], whereas all five ancestral 6p25 segmental duplications and deletions are non-recurrent (Figure 2.1A: I-V). This indication that 6p25 may differ from previously studied genomic regions, is supported by the substantially broader than previously described segmental anomaly size spectrum (~30 – 1220 kb) observed with these CNVs. The second

feature relates to the sequences that mediate the rearrangements, which have generally been low copy repeats (LCRs), comprising elements some 10-500 kb in size that share ≥ 95% sequence homology [42]. Although no LCRs are present in this region, convincing evidence that blocks of repetitive sequence contribute to the rearrangements is provided by location of breakpoints within (#1-2, 8) or adjacent (#3-6, 7) to repeat elements in 4 of the 5 ancestral rearrangements, that are themselves located within large repeat blocks. In this context, it is noteworthy that in #3-6 (Figure 2.2) the adjacent repeat block is extensive (13.8 kb), meeting the LCR size criterion. Furthermore, in pedigrees #7 and 8 (Figure 2.3 & 2.4), location of members of the same repeat element class at or adjacent to each breakpoint, provides putative recombination substrates.

In addition, all five ancestral rearrangements exhibit NHEJ with base pair insertion or deletion at the junction. Multiple rearrangements exhibiting both NAHR and NHEJ have only been previously reported in a subset of X-chromosomal duplications (Pelizaeus-Merzbacher disease, MIM 312080) [54]. Concurrent involvement of NAHR and NHEJ on an autosome, and particularly with both segmental duplication and deletion, convincingly demonstrates that a greater proportion of rearrangements use mixed mechanisms. The differing degrees of NAHR and NHEJ evident in the five rearrangements, indicate that these represent components of a mechanistic spectrum (Figure 2.8).

Figure 2.8 Mechanistic spectrum observed in the five ancestral segmental rearrangements.

The degree of NAHR and or NHEJ is depicted by each rearrangement's relative position. The repeat elements at or adjacent to the breakpoints are shown as underlined or italicized text, respectively.


NAHR
Mechanistic Spectrum of NAHR and NHEJ
Duplication: #1-2 (LINE / SINE)
Duplication: #7 (SINE / SINE)
Duplication: #3-6 (DNA repeat / SINE)
Deletion: #8 (Simple repeat / Simple Repeat)
Deletion: #9 (No repeat elements)
NHEJ

The architecture of rearrangement #8 (Figure 2.4) is unprecedented, with a 60-fold larger insert at the junction than previously observed in NHEJ (6 bp) [54], as well as features of NAHR (breakpoints within GTGn repeats). The insert's size (367bp), identical motifs, and particularly homology to sequence adjacent to both breakpoints (Figure 2.4A) are incompatible with NAHR and NHEJ being the sole mechanisms involved. Indeed, existence of DNA polymerase α pause and Topo-isomerase I consensus cleavage sites at, or adjacent to, both breakpoints (Figure 2.4B) accord with a more complex mechanism, notably the recently described replication based mechanism observed in two non-recurrent Pelizaeus-Merzbacher disease (PMD) cases [40]. The DNA polymerase pause sites and extensive (GTG)n at the breakpoints represent recognized impediments to polymerase progression [55, 56], capable of causing the same successive replication fork stalling that characterized the PMD cases [40]. Stalled forks are recombinogenic [57], in part due to the potential for the increased length of single strand DNA to lead to double strand break and subsequent template switching [40].

Hence the most parsimonious explanation for the observed architecture of motif M1 is recurrent stalling followed by re-initiation of strand synthesis ahead of the sequence that induced the fork to stall. This is in agreement with the extensive homology between M1 and the three sequence blocks (labeled I - III, Figure 2.4B) that straddle the centromeric breakpoint, which lie in the correct order and orientation to yield this 128 bp motif portion. The imperfect sequence identity at the

junctions of blocks 'I' and 'II', and 'II' and 'III' is explicable by error-prone repair of the resultant single strand gaps using either the sister chromatid, or other sequence, as template [58]. The key differences with the Fork Stalling and Template Switching Mechanism [40] concern the absence of evidence of template switching in #8 and the higher order symmetry of the identical motifs. The latter is compatible with a double strand break at the telomeric breakpoint – the location of a topoisomerase I cleavage site (CAT) - followed by single strand dependent synthesis using the antisense strand of M1 as template to generate M2. Such findings, albeit based upon a single complex rearrangement, accord with the central tenet of the replication-based mechanism [40], whilst illustrating potential for additional intricacies and demonstrating involvement of this process in segmental deletion in addition to the previously described two duplications [40].

The unique architecture of 6p25 amongst regions prone to segmental duplication and deletion[23-30] with three tandemly arranged paralogs (*FOXC1, FOXF2* and *FOXQ1*), heightens interest in these rearrangements. CNVs in regions containing paralogous triplets cause both human [colour blindness: opsin cluster, Xq28 (MIM 303800)] [59, 60] and canine phenotypes [ridge and dermal sinus: *FGF3/4/19* cluster, chr. 18] [61]. Since homology between the parologous forkhead domains is at a level sufficient to promote strand exchange [62, 63], their presence is compatible with a role in CNV formation, possibly by promoting misalignment or stabilizing intermediate structures. CNV prevalence is increased in genomic regions

characterized by segmental duplication and the available evidence indicate that the 6p25 (*FOXC1*, *FOXF2*, *FOXQ1*) and 16q24 clusters (*FOXC2*, *FOXF1*, *FOXL1*) (MIM 602402, 601089, 603252) were generated by ancestral segmental duplication events [21, 64]. This raises the intriguing possibility that the segmental duplications that generated the 6p25 forkhead triplet may predispose to subsequent copy number variation.

Defining the breakpoint architecture of each of the five ancestral segmental rearrangements, a particular strength of this study, revealed a marked discrepancy between array CGH-predicted and actual breakpoints in #7 (Figure 2.3). This novel observation highlights potential for sequence variation to confound CNV analysis with particular clinical relevance due to increasing transition of array CGH from the research laboratory into clinical testing. We attribute this finding to masking by paralogous test or reference sequence, since array CGH copy number data reflect the overall level of hybridization to individual probes. In view of the potential for under-ascertainment of smaller CNVs, there will be a need to evaluate the technique's sensitivity and specificity prior to widespread diagnostic testing.

In summary, our findings illustrate the value of undertaking the extensive experiments necessary to characterize individual CNVs at the base pair level, providing evidence of a novel spectrum of mechanisms that mediate 6p25 rearrangements. This combination of recombination, DNA repair and replication-

based mechanisms, together with the CNV size spectrum and 100% non-recurrence rate, highlight unique features of this genomic region. Use of increasingly higher resolution techniques to interrogate 6p25 has detected progressively smaller CNVs, providing translational opportunities for determining the molecular basis of an appreciable proportion of pediatric glaucoma. Although CNVs smaller than the current chromosome 6p25 ~30 kb minimum are present genome-wide, their prevalence is imperfectly defined due to most current arrays' ≥5 kb probe densities. If our hypothesis is correct that formation of small structural variants may be mediated by smaller sequence elements than larger CNVs (as in 6p25), analysis of regions containing multiple paralogous dosage sensitive genes that exhibit a particularly broad spectrum of rearrangements, may be especially fruitful in determining novel causes of copy number variation.

Web Resources

The accession number and URLs for data presented herein are as follows:

Ensembl Genome Browser, http://www.ensembl.org/index.html (for version 36)

NCBI BLAST, http://www.ncbi.nlm.nih.gov/blast/

NCBI Entrez, http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/

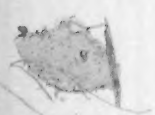
UCSC Genome Browser, http://genome.ucsc.edu/

UniGene, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene

Pip Maker, http://pipmaker.bx.psu.edu/pipmaker/

Repeat Masker, http://repeatmasker.org/cgi-bin/WEBRepeatMasker

Palindrome, http://bioweb.pasteur.fr/seqanal/interfaces/palindrome.html



Bibliography

1. Tuzun, E., et al., *Fine-scale structural variation of the human genome.* Nat Genet, 2005. **37**(7): p. 727-32.

2. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

3. Perry, G.H., et al., *Diet and the evolution of human amylase gene copy number variation.* Nat Genet, 2007. **39**(10): p. 1256-60.

4. Hollox, E.J., J.A. Armour, and J.C. Barber, *Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster.* Am J Hum Genet, 2003. **73**(3): p. 591-600.

5. Moessner, R., et al., *Contribution of SHANK3 mutations to autism spectrum disorder.* Am J Hum Genet, 2007. **81**(6): p. 1289-97.

6. Yang, Y., et al., *Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans.* Am J Hum Genet, 2007. **80**(6): p. 1037-54.

7. Sebat, J., et al., *Strong association of de novo copy number mutations with autism.* Science, 2007. **316**(5823): p. 445-9.

8. Hollox, E.J., et al., *Psoriasis is associated with increased beta-defensin genomic copy number.* Nat Genet, 2008. **40**(1): p. 23-5.

9. Scherer, S.W., et al., *Challenges and standards in integrating surveys of structural variation.* Nat Genet, 2007. **39**(7 Suppl): p. S7-15.

10. Kleinjan, D.A. and V. van Heyningen, *Long-range control of gene expression: emerging mechanisms and disruption in disease.* Am J Hum Genet, 2005. **76**(1): p. 8-32.

11. Schedl, A., et al., *Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities.* Cell, 1996. **86**(1): p. 71-82.

12. Nishimura, D.Y., et al., *A spectrum of FOXC1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye.* Am J Hum Genet, 2001. **68**(2): p. 364-72.

13. Lehmann, O.J., et al., *Chromosomal duplication involving the forkhead transcription factor gene FOXC1 causes iris hypoplasia and glaucoma.* Am J Hum Genet, 2000. **67**(5): p. 1129-35.

14. Fantes, J., et al., *Mutations in SOX2 cause anophthalmia.* Nat Genet, 2003. **33**(4): p. 461-3.

15. Asai-Coakwell, M., et al., *GDF6, a Novel Locus for a Spectrum of Ocular Developmental Anomalies.* Am J Hum Genet, 2007. **80**(2): p. 306-15.

16. Kume, T., et al., *The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus.* Cell, 1998. **93**(6): p. 985-96.

17. Lehmann, O.J., et al., *Ocular developmental abnormalities and glaucoma associated with interstitial 6p25 duplications and deletions.* Invest Ophthalmol Vis Sci, 2002. **43**(6): p. 1843-9.

18. Maclean, K., et al., *Axenfeld-Rieger malformation and distinctive facial features: Clues to a recognizable 6p25 microdeletion syndrome.* Am J Med Genet A, 2005. **132**(4): p. 381-5.

19. Gould, D.B., et al., *Phenotypic and molecular assessment of seven patients with 6p25 deletion syndrome: relevance to ocular dysgenesis and hearing impairment.* BMC Med Genet, 2004. **5**: p. 17.

20. Strungaru, M.H., I. Dinu, and M.A. Walter, *Genotype-phenotype correlations in Axenfeld-Rieger malformation and glaucoma patients with FOXC1 and PITX2 mutations.* Invest Ophthalmol Vis Sci, 2007. **48**(1): p. 228-37.

21. Lehmann, O.J., et al., *Fox's in development and disease.* Trends Genet, 2003. **19**(6): p. 339-44.

22. Lehmann, O.J., et al., *Novel anterior segment phenotypes resulting from forkhead gene alterations: evidence for cross-species conservation of function.* Invest Ophthalmol Vis Sci, 2003. **44**(6): p. 2627-33.

23. Lupski, J.R., et al., *DNA duplication associated with Charcot-Marie-Tooth disease type 1A.* Cell, 1991. **66**(2): p. 219-32.

24. Chance, P.F., et al., *DNA deletion associated with hereditary neuropathy with liability to pressure palsies.* Cell, 1993. **72**(1): p. 143-51.

25. Potocki, L., et al., *Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion.* Nat Genet, 2000. **24**(1): p. 84-7.

26. Yobb, T.M., et al., *Microduplication and triplication of 22q11.2: a highly variable syndrome.* Am J Hum Genet, 2005. **76**(5): p. 865-76.

27. Edelmann, L., et al., *A common molecular basis for rearrangement disorders on chromosome 22q11.* Hum Mol Genet, 1999. **8**(7): p. 1157-67.

28. Somerville, M.J., et al., *Severe expressive-language delay related to duplication of the Williams-Beuren locus.* N Engl J Med, 2005. **353**(16): p. 1694-701.

29. Long, F.L., et al., *Triplication of 15q11-q13 with inv dup(15) in a female with developmental delay.* J Med Genet, 1998. **35**(5): p. 425-8.

30. Ledbetter, D.H., et al., *Deletions of chromosome 15 as a cause of the Prader-Willi syndrome.* N Engl J Med, 1981. **304**(6): p. 325-9.

31. Beutler, E., T. Gelbart, and W. Kuhl, *Interference of heparin with the polymerase chain reaction.* Biotechniques, 1990. **9**(2): p. 166.

32. Tsai, M., et al., *Detection of mouse mast cell-associated protease mRNA. Heparinase treatment greatly improves RT-PCR of tissues containing mast cell heparin.* Am J Pathol, 1995. **146**(2): p. 335-43.

33. Mears, A.J., et al., *Autosomal dominant iridogoniodysgenesis anomaly maps to 6p25.* Am J Hum Genet, 1996. **59**(6): p. 1321-7.

34. Walter, M.A., et al., *Autosomal-dominant iridogoniodysgenesis and Axenfeld-Rieger syndrome are genetically distinct.* Ophthalmology, 1996. **103**(11): p. 1907-15.

35. Mears, A.J., et al., *Mutations of the forkhead/winged-helix gene, FKHL7, in patients with Axenfeld-Rieger anomaly.* Am J Hum Genet, 1998. **63**(5): p. 1316-28.

36. Dhami, P., et al., *Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome.* Am J Hum Genet, 2005. **76**(5): p. 750-62.

37. Selzer, R.R., et al., *Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH.* Genes Chromosomes Cancer, 2005. **44**(3): p. 305-19.

38. Meese, E.U., et al., *Alu-PCR: characterization of a chromosome 6-specific hybrid mapping panel and cloning of chromosome-specific markers.* Genomics, 1992. **12**(3): p. 549-54.

39. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.

40. Lee, J.A., C.M. Carvalho, and J.R. Lupski, *A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders.* Cell, 2007. **131**(7): p. 1235-47.

41. Kanaar, R., J.H. Hoeijmakers, and D.C. van Gent, *Molecular mechanisms of DNA double strand break repair.* Trends Cell Biol, 1998. **8**(12): p. 483-9.

42. Stankiewicz, P. and J.R. Lupski, *Genome architecture, rearrangements and genomic disorders.* Trends Genet, 2002. **18**(2): p. 74-82.

43. Shaw, C.J. and J.R. Lupski, *Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease.* Hum Mol Genet, 2004. **13 Spec No 1**: p. R57-64.

44. Lupski, J.R. and P. Stankiewicz, *Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes.* PLoS Genet, 2005. **1**(6): p. e49.

45. Raghavan, S.C., J. Tong, and M.R. Lieber, *Hybrid joint formation in human V(D)J recombination requires nonhomologous DNA end joining.* DNA Repair (Amst), 2006. **5**(2): p. 278-85.

46. Schwartz, M., et al., *Homologous recombination and nonhomologous end-joining repair pathways regulate fragile site stability.* Genes Dev, 2005. **19**(22): p. 2715-26.

47. Sonoda, E., et al., *Differential usage of non-homologous end-joining and homologous recombination in double strand break repair.* DNA Repair (Amst), 2006. **5**(9-10): p. 1021-9.

48. Lieber, M.R., et al., *Mechanism and regulation of human non-homologous DNA end-joining.* Nat Rev Mol Cell Biol, 2003. **4**(9): p. 712-20.

49. Tsukamoto, Y. and H. Ikeda, *Double-strand break repair mediated by DNA end-joining.* Genes Cells, 1998. **3**(3): p. 135-44.

50. Roth, D.B., T.N. Porter, and J.H. Wilson, *Mechanisms of nonhomologous recombination in mammalian cells.* Mol Cell Biol, 1985. **5**(10): p. 2599-607.

51. Pentao, L., et al., *Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit.* Nat Genet, 1992. **2**(4): p. 292-300.

52. Juyal, R.C., et al., *Molecular analyses of 17p11.2 deletions in 62 Smith-Magenis syndrome patients.* Am J Hum Genet, 1996. **58**(5): p. 998-1007.

53. Park, S.S., et al., *Structure and evolution of the Smith-Magenis syndrome repeat gene clusters, SMS-REPs.* Genome Res, 2002. **12**(5): p. 729-38.

54. Woodward, K.J., et al., *Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination.* Am J Hum Genet, 2005. **77**(6): p. 966-87.

55. Mirkin, E.V. and S.M. Mirkin, *Replication fork stalling at natural impediments.* Microbiol Mol Biol Rev, 2007. **71**(1): p. 13-35.

56. Fouche, N., et al., *Replication fork regression in repetitive DNAs.* Nucleic Acids Res, 2006. **34**(20): p. 6044-50.

57. Lambert, S., et al., *Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier.* Cell, 2005. **121**(5): p. 689-702.

58. Langston, L.D. and M. O'Donnell, *DNA replication: keep moving and don't mind the gap.* Mol Cell, 2006. **23**(2): p. 155-60.

59. Nathans, J., et al., *Molecular genetics of human visual pigments.* Annu Rev Genet, 1992. **26**: p. 403-24.

60. Deeb, S.S. and A.G. Motulsky, *Molecular genetics of human color vision.* Behav Genet, 1996. **26**(3): p. 195-207.

61. Salmon Hillbertz, N.H., et al., *Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs.* Nat Genet, 2007. **39**(11): p. 1318-20.

62. Waldman, A.S. and R.M. Liskay, *Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology.* Mol Cell Biol, 1988. **8**(12): p. 5350-7.

63. Lam, K.W. and A.J. Jeffreys, *Processes of copy-number change in human DNA: the dynamics of {alpha}-globin gene deletion.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 8921-7.

64. Carlsson, P. and M. Mahlapuu, *Forkhead transcription factors: key players in development and metabolism.* Dev Biol, 2002. **250**(1): p. 1-23.

Chapter 3

Towards identifying novel genetic factors of ARS

Introduction:

The genetic heterogeneity of ARS is apparent from the fact that five genetic loci (6p25 [1, 2], 4q25 [3], 16q24 [4], 13q14 [5] and 11p13 [6]) have been associated with this disorder. However, currently only two genes, *FOXC1* (6p25) and *PITX2* (4q25), have been identified where both mutations and CNVs result in ARS. Estimation that ~60% of patients with ARS have unknown genetic background [7], suggest further genetic heterogeneity in ARS. Since ~50% of ARS patients develop glaucoma, one of the leading causes of irreversible blindness worldwide, a systematic approach to identify the unknown genetic factors of ARS is warranted.

An ARS panel of 32 patients was screened for both mutations and copy number variations in *FOXC1,* and for only mutations in *PITX2*. While screening for mutations was done primarily by sequencing, and if possible, supplemented by restriction enzyme (R.E.) digestion, CNVs were screened for by quantitative or real time PCR and microsatellite marker genotyping.

Materials and methods:

PCR:

Four primer pairs, separately designed for *FOXC1* and *PITX2* genes, were used for PCR based amplification of these genes (Table 3.1). The primers were designed using Primer 3 (V.0.4.0). Different reagents and programs of cycling temperature were used for amplification of *FOXC1* and *PITX2*.

a) *FOXC1*-

Cycling temperature: (1) 95° C- 2 min, (2) 95° C- 30sec, (3) 58° C- 1 min, (4) 72° C- 1 min, (5) Repeat (2-4) 34 times, (6) 72° C- 5 min, (7) Hold at 15°C, (8) End.

Chemical reagents: Each reaction, carried out in a total of 25µl, had a final PCR buffer concentration of 1X (10X is 200mM Tris, pH 8.4, 500 mM KCl; Invitrogen ®), dNTP concentrations of 0.2mM (Invitrogen®), $MgCl_2$ concentration of 1.5mM (Invitrogen®), glycerol concentration of 10% v/v (Anachemia®), final formamide concentration of 6% v/v (Sigma®) and 0.025 U of Taq polymerase (New England BioLabs® and Invitrogen ®).

b) *PITX2*-

Cycling temperature: (1) 95° C- 5 min, (2) 95° C- 30sec, (3) 60° C- 1 min, (4) 72° C- 1 min, (5) Repeat (2-4) 29 times, (6) 72° C- 5 min, (7) Hold at 15°C, (8) End.

Chemical reagents: Each reaction, carried out in a total of 25µl, had a final PCR buffer concentration of 1X (10X is 200mM Tris, pH 8.4, 500 mM KCL; Invitrogen ®), dNTP concentrations of 0.2mM (Invitrogen®), $MgCl_2$ concentration of 2.0mM (Invitrogen®), and 0.025 U of Taq polymerase (New England BioLabs® and Invitrogen ®).

Amplified PCR products were validated by gel electrophoresis- ethidium bromide stained 1-1.5% agarose gel. After validation, the PCR products were purified (Sigma® Spin Post –Reaction Clean-Up Columns) prior to sequencing.

Sequencing:

The sequencing reactions were done with the Applied Biosystems BigDye® Terminator v3.1 Cycle sequencing kit and the unincorporated dyes were removed with ethanol. Sequencing was run on the Applied Biosystems (ABI) 3130/XL (The Applied Genomic Centre, Alberta, Canada) or 3737/XL (CHUL Research Center, Quebec, Canada) Genetic Analyzer with the same primer pairs used to amplify *FOXC1* and *PITX2* (Table 3.1).

Restriction enzyme digestion:

Some of the mutations identified in *FOXC1* and *PITX2* were validated with restriction enzyme (R.E.) digestion. Two restriction enzymes, *BasaHI* (New England BioLabs®; recognition sequence-GRCGYC) and *Sfc1* (New England

BioLabs®; recognition sequence- CTRYAG), were used for validating mutations in *PITX2* and *FOXC1* respectively. 10µl of PCR products (at concentration of ~100ng/µl) was incubated with 1µl of restriction enzymes (20,000 U/ml) at 37°C for 5 hours. The digested products were analysed with ethidium bromide stained 1% agarose gel electrophoresis.

Real time-PCR:

To identify *FOXC1* CNV in the ARS panel, ABI Taqman® real time PCR was adopted. The primers and probes (Table 2.2) for the Taqman® real time PCR were designed using ABI Primer Express® Software v.3.0. The real time PCR reaction was carried out in Applied Biosystem® AB-7900-HT Sequence Detection System.

Cycling temperature: (1) 50° C- 2 min, (2) 95° C- 10 min, (3) 95° C- 15 sec, (4) 60° C- 1 min, (5) Repeat (3-4) 39 times, (6) Hold at 4°C, (7) End.

Chemical reagents: Each reaction, carried out in a total of 15µl, had a final Taqman® buffer concentration of 50% v/v (Applied BioSystems®) and final probe concentration of 1.33% v/v at 50ng/ µl (Applied BioSystems®).

Microsatellite marker genotyping:

Three microsatellite markers, *D6S967*, *BA121* and *FF2*, designed in the vicinity of *FOXC1*, were used to screen the ARS panel for putative CNVs (Table 2.3 and Figure 3.2). The primers for *BA121* and *FF2* were designed using Primer 3 (v. 0.4.0)

while the primer pair for D6S967 was adopted from NCBI. The primers (either forward or reverse) were fluorescently labeled by Applied Biosystems®. The reaction condition for microsatellite marker genotyping was the same as that used for PCR of *PITX2*.

Table 3.1. Primer pairs used for *FOXC1* and *PITX2* amplifications.

FOXC1	Forward	Reverse	Annealing Temp (°C)
F1	GAAGCTGCGCCGCGAGTTCCTG	GAAGCGGTCCATGATGAACTGGTAG	58
F2	CGGACAAGAAGATCACCCTGAACG	CGTACCGTTCTCGGTCTTGATGTC	58
F3	AGGCTGCACCTCAAGGAGCCGC	GTGACCGGAGGCAGAGAGTAGTC	58
F4	TACCACTGCAACCTGCAAGC	GGGTTCGATTTAGTTCGGCT	58

PITX2	Forward	Reverse	Annealing Temp (°C)
P1	TAACCTCTGGGCACTTTTGC	GCGATTTGGTTCTGATTTCC	60
P2	TGGGTCTTTGCTCTTTGTCC	CCAGAGGCGGAGTGTCTAAG	60
P3	AGCTCTTCCACGGCTTCTG	GGGAACTGTAATCTCGCAACC	60
P4	AGTGCGCTAGCGTGTGTG	GAGCTCTCTCTTTGATTCAGTGG	60

Results:

Three novel mutations in *FOXC1* and one novel mutation in *PITX2* were identified. Of the three mutations in *FOXC1*, two [(Y81D) and (W122G)] lie within the forkhead domain while the third (W5Stop) lies in the N-terminal transactivation domain (Figure 3.1) The mutation (Y81D) was validated by restriction enzyme digestion (Figure 3.1C). Restriction enzymes recognizing the sequence changes of the other two mutations were not available commercially. In *PITX2*, the only identified mutation (W213*fs*) results in frame shift of the PITX2 coding sequence (Figure 3.2A-C). *In-silico* analysis predicts the prematurely terminated PITX2 protein to have unique 24 amino acids after codon 213 and is expected to be short of the Homeo-OAR domian. Restriction enzyme digestion was adopted to validate this mutation (Figure 3.2D). Mutations in FOXC1 and PITX2 identified in previous and present studies are summarized in Tables 3.2 and 3.3 respectively.

In addition to mutations, the screening of the ARS panel with the microsatellite marker *D6S967* revealed a small (~329bp) putative copy number variant (duplication) region, approximately 127kb telomeric to *FOXC1*, in proband 11 (Figure 3.3). However, customized 6p25 oligonucleotide array on proband 11 suggested no such putative CNVs (Figure 3.3). Screening of the ARS panel with two other microsatellite markers (*BA121* and *FF2*) did not identify any CNVs.

Figure 3.1 *FOXC1* mutations

A. Chromatograms displaying heterozygous *FOXC1* mutations (arrows) identified in probands (P) 11-13.
B. Effect of the mutations in the protein level. Location (arrows) and amino acid changes have been shown.
C. Overview of the *FOXC1* alignments in different species. The mutated nucleotides are highlighted in red.
D. Restriction enzyme digestion to confirm the mutation Y81D. The mutation abolishes the cutting site of *SfcI*. Hence, only in P12, in addition to the normal digested fragments (530 and 480bp), the original PCR product (1072bp) remains undigested. R.E. digestion confirms the mutation to be heterozygous.

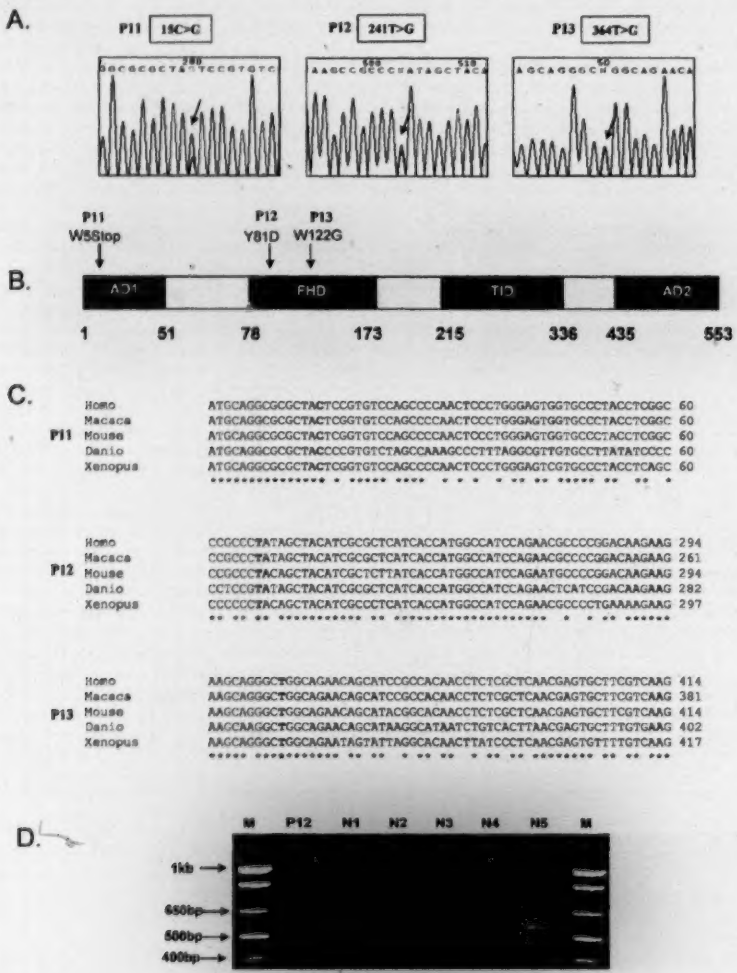


Figure 3.2 *PITX2* mutation

A. Chromatogram displaying the mutation 649delG (arrow) in *PITX2*.
B. In-silico prediction of the effect of the mutation at the protein level. The prematurely terminated PITX2 has 24 unique amino acids (red) after the codon 213 (red open rectangle).
C. Conservation status of the mutated nucleotide in different species.
D. Restriction digestion with BsaHI to confirm the mutation Trp213*fs*. The mutation introduces an additional cutting site for the R.E. As such, only in proband 14, harboring the mutation, an additional 415bp band was present. R.E. digestion confirms the mutation to be heterozygous.

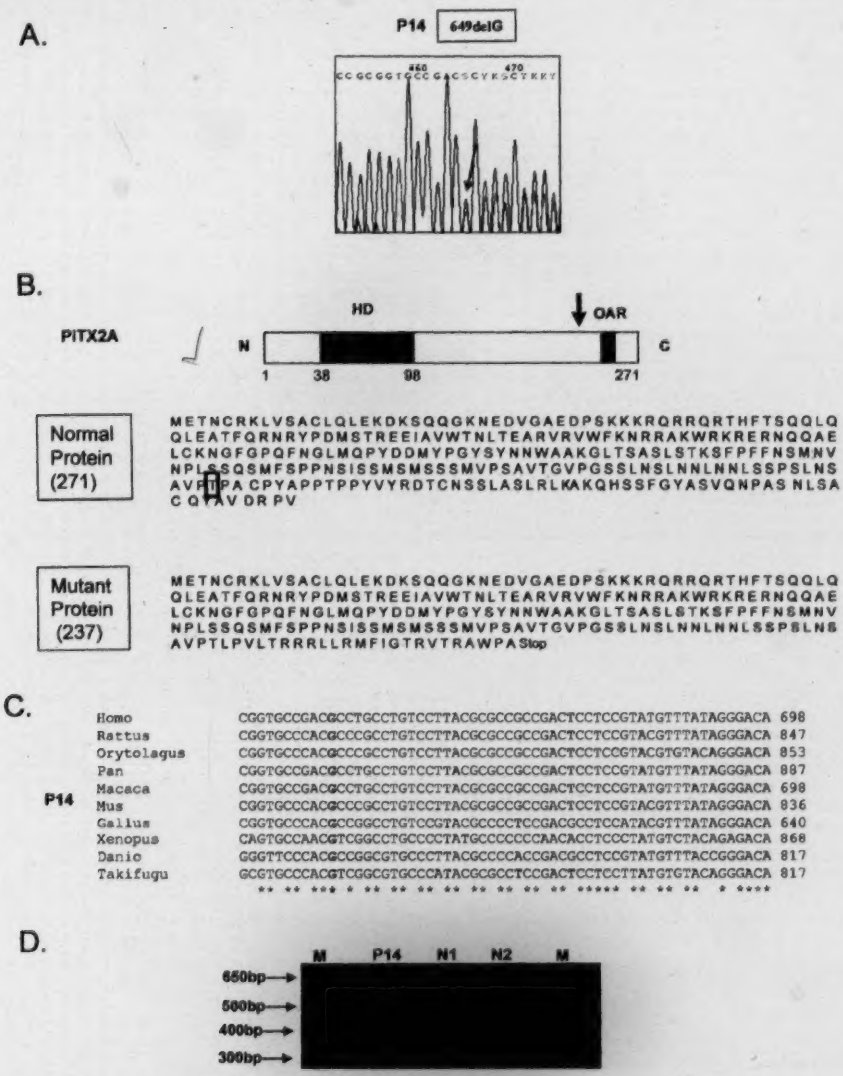


Figure 3.3: Montage illustrating the relative positions of the different microsatellite markers and the 6p25 Fox triplet.

A. Chromosome 6.
B. Chromosome 6p25.
C. The relative positions of the microsatellite markers and Fox triplet (boxed).

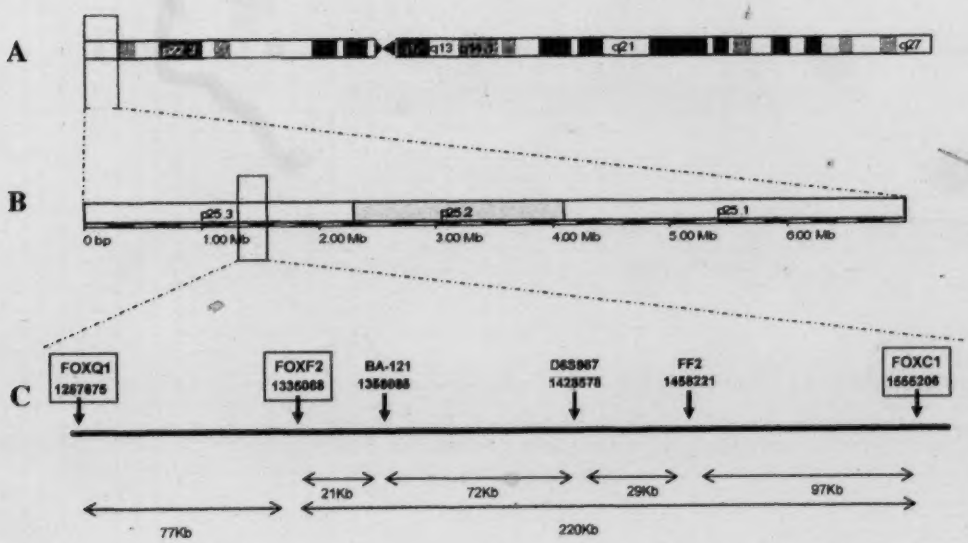


Figure 3.4: Montage illustrating the discrepancy between microsatellite marker genotyping and oligonucleotide array result.

A. In proband 15, height of one allele (denoted by black arrow), corresponding to marker *D6S967*, is twice that of the other, which is indicative of duplication. In a normal individual, the peaks of the alleles are at equal heights.

B. Customized oligonucleotide array in proband 15 showing 3Mb of 6p25. The view in and around *D6S967* is magnified. The $\log_2$ ratio of the probes encompassed by *D6S967* (blue shaded) is not indicative of any CNV. Note the genomic location of *D6S967* is from 1428578-1428900 on chromosome 6.

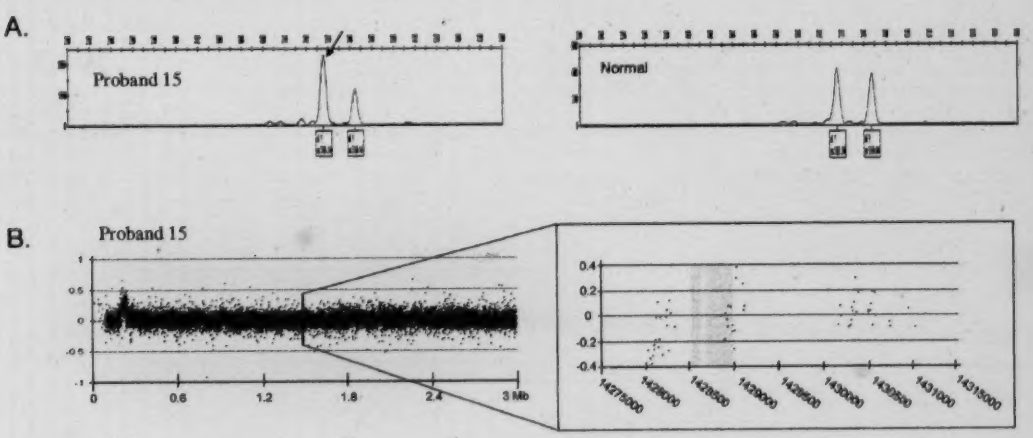


Table 3.2: *FOXC1* mutations identified in previous and present studies. Adopted from [8]

Mutations in *FOXC1* identified in previous studies.

	Nucleotide Change	Protein Change	References
1	93-102del	Nonsense	Mears *et al.* [1]
2	245G>C	Missense (S82T)	
3	261C>G	Missense (I87M)	
4	153-163del	Nonsense	Nishimura *et al.* [2]
5	335T>C	Missense (F112S)	
6	378C>G	Missense (I126M)	
7	392C>T	Missense (S131L)	
8	210delG	Nonsense	Swiderski *et al.* [9]
9	67C>T	Nonsense (Q23Stop)	Mirzayans *et al.* [10]
10	99-108del	Nonsense	Nishimura *et al.* [11]
11	116-123del	Nonsense	
12	1512del	Nonsense	
13	265insC	Nonsense	
14	26-47ins	Nonsense	
15	236C>T	Missense (P79L)	
16	286insG	Nonsense	Kawase *et al.* [12]
17	272T>G	Missense (I91S)	
18	380G>A	Missense (R127H)	
19	235C>A	Missense (P79T)	Suzuki *et al.* [13]
20	482T>A	Missense (M161K)	Panicker *et al.* [14]
21	255GCTT	Missense (L86F)	Saleem *et al.* [15]
22	4C>T	Nonsense (Q2Stop)	Komatireddy *et al.* [16]
23	367C>T	Nonsense (Q123Stop)	
24	272T>C	Missense (I91T)	Mortemousque *et al.* [17]
25	494G>C	Missense (G165R)	Murphy *et al.* [18]
26	506G>C	Missense (R169P)	
27	738delG	Nonsense	Weisschuh *et al.*[8]
28	1511delT	Nonsense	
29	143C>A	Nonsense (S48X)	
30	236C>G	Missense (P79R)	
31	339T>C	Missense (Y115S)	
32	446G>A	Missense (G149D)	
33	481A>G	Missense (M161V)	
34	388C>T	Missense (L130F)	Ito *et al.* [19]
35	253G>C	Missense (A85P)	Fuse *et al.*[20]
36	437-453del	Nonsense	

Mutations in *FOXC1* identified in the present study.

	Nucleotide Change	Protein Change	Proband
1	15C>G	Nonsense (W5Stop)	P11
2	241T>G	Missense (Y81D)	P12
3	364G>T	Missense (W122G)	P13

Table 3.3: *PITX2* mutations identified in previous and present studies. Adopted from [8]

Mutations in *PITX2* identified in previous studies.

	Nucleotide Change	Protein Change	References
1	744T>A	Missense (L54Q)	Semina *et al.* [3]
2	785A>C	Missense (T68P)	
3	855G>C	Missense (R91P)	
4	981G>A	Nonsense (W133X)	
5	Exon 3(-11)A>G	Splice site	
6	Exon 3(+5)G>C	Splice site	
7	789G>A	Missense (R69H)	Kulak *et al.*[21]
8	833C>T	Missense (R84W)	Alward *et al.*[22]
9	Exon 3(-2)A>T	Splice site	Doward *et al.* [23]
10	845A>G	Missense (K88E)	Perveen *et al.*[24]
11	851C>T	Missense (R90C)	
12	Exon 2(-1)G>C	Splice site	
13	1083insC	Nonsense	
14	868-869delAA	Nonsense	
15	939delA	Nonsense	
16	1235-1236TA>AAG	Nonsense	
17	830G>C	Missense (V83L)	Priston *et al.* [25]
18	713-733dupl	Nonsense	
19	1272delG	Nonsense	Borges *et al.* [26]
20	774C>T	Missense (P64L)	Philips *et al.* [27]
21	852G>C	Missense (R90P)	
22	896C>G	Missense (L105V)	
23	906A>C	Missense (N108T)	
24	717-720delACTT	Nonsense	Wang *et al.*[28]
25	1261delT	Nonsense	Brooks *et al.*[29]
26	697delG	Nonsense	Lines *et al.* [30]
27	Exon 3(-1)G>T	Nonsense	
28	998delC	Nonsense	
29	959delC	Nonsense	Saadi *et al.* [31]
30	710C>T	Missense (R43W)	Idrees *et al.* [32]
31	774C>G	Missense (P64R)	Weisschuh *et al.* [8]
32	753C>G	Missense (F58L)	Vieira *et al.* [33]
33	742G>T	Nonsense (E55Stop)	
34	942C>A	Nonsense (Y121Stop)	
35	1251insCGACTCCT	Nonsense	

Mutation in *PITX2* identified in the present study.

	Nucleotide Change	Protein Change	Proband
1	649delG	Nonsense	P14

Discussion:

The pathological consequences of the *FOXC1* and *PITX2* mutations await functional characterization (*in vitro* and *in vivo).* However, a preliminary outcome can be drawn based on the location and nature of these mutations. The missense mutations, Y81D and W122G, are located in the forkhead domain of *FOXC1* (Figure 3.1) and a previous study has shown that missense mutations in the FHD of *FOXC1* results in reduced transactivation and/or altered DNA binding activity of FOXC1 [34]. As such, it can be hypothesized that these mutations will manifest ARS by either reducing the transactivation or by altering the binding specificity of FOXC1. The other mutation, W5Stop (Figure 3.1), will result in premature termination of FOXC1. The mutant FOXC1 will have only five amino acid residues. Such premature termination will likely be equivalent to haplo-insufficiency that results from *FOXC1* deletion. Hence, patients with W5Stop might have similar pathological outcomes to that of patients with heterozygous *FOXC1* deletion ($FOXC^{+/-}$). In *PITX2*, the only identified mutation (Trp231*fs*) results in generation of PITX2 protein that lacks the Homeo_OAR domain. Such data highlight the importance of the Homeo_OAR domain for proper functioning of PITX2. It is noteworthy that all the mutations in *FOXC1* and *PITX2* affect residues that are conserved amongst different species (Figure 3.1 and 3.2). In summary, the screening of 32 ARS patients identified novel *FOXC1* or *PITX2* mutations in only 4 individuals. Such data, taken

together with results from previous study [7], support the existence of further genetic heterogeneity in ARS.

Another interesting finding of the current study is the discrepancy between microsatellite marker analyses and customized oligonucleotide array CGH results in proband 15 (Figure 3.3). The microsatellite marker *D6S967* identified a putative CNV, approximately 127 kb telomeric to *FOXC1* (Figure 3.3). However, customized oligonucleotide array CGH with average probe density of 100bp, revealed no such CNV in the region corresponding to *D6S967*. This discrepancy could be attributable to different caveats that are associated with either oligonucleotide array or microsatellite marker genotyping. As described in Chapter 1 (General Introduction), the existence of paralogous regions in the test or reference DNA; decrease in signal to noise ratio from increased probe density and limitations in the probe design, could contribute towards erroneous oligonucleotide array CGH results. On the contrary, the high probe density (avg 100bp) and isothermal probe design in customized oligonucleotide array CGH (NimbleGen®) are expected to provide high sensitivity in CNV detection.

In microsatellite marker genotyping, the availability of the two alleles, simultaneously and efficiently, for primer binding during the exponential phase of PCR is crucial for the accuracy in the results. It is possible that even when none of the alleles have CNVs, the relative accessibility of one allele to the other for primer

binding could produce false positive results. However, previous studies have validated *D6S967* positive samples to have 6p25 CNVs by either FISH [35, 36] or CGH [37]. Hence, it is unlikely that *D6S967* would produce a false positive result in proband 15. Furthermore, microsatellite marker analysis of proband 15 with *D6S967* was repeated thrice with no difference in the result. As such, at present the explanation for the apparent discrepancy between microsatellite marker genotyping and customized oligonucleotide array CGH results is not absolute and can not be resolved.

Bibliography

1. Mears, A.J., et al., *Mutations of the forkhead/winged-helix gene, FKHL7, in patients with Axenfeld-Rieger anomaly.* Am J Hum Genet, 1998. **63**(5): p. 1316-28.

2. Nishimura, D.Y., et al., *The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25.* Nat Genet, 1998. **19**(2): p. 140-7.

3. Semina, E.V., et al., *Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome.* Nat Genet, 1996. **14**(4): p. 392-9.

4. Werner, W., et al., *A small deletion of 16q23.1-->16q24.2 [del(16)(q23.1q24.2).ish del(16)(q23.1q24.2)(D16S395+, D16S348-, P5432+)] in a boy with iris coloboma and minor anomalies.* Am J Med Genet, 1997. **70**(4): p. 371-6.

5. Phillips, J.C., et al., *A second locus for Rieger syndrome maps to chromosome 13q14.* Am J Hum Genet, 1996. **59**(3): p. 613-9.

6. Riise, R., K. Storhaug, and K. Brondum-Nielsen, *Rieger syndrome is associated with PAX6 deletion.* Acta Ophthalmol Scand, 2001. **79**(2): p. 201-3.

7. Strungaru, M.H., I. Dinu, and M.A. Walter, *Genotype-phenotype correlations in Axenfeld-Rieger malformation and glaucoma patients with FOXC1 and PITX2 mutations.* Invest Ophthalmol Vis Sci, 2007. **48**(1): p. 228-37.

8. Weisschuh, N., et al., *Novel mutations of FOXC1 and PITX2 in patients with Axenfeld-Rieger malformations.* Invest Ophthalmol Vis Sci, 2006. **47**(9): p. 3846-52.

9. Swiderski, R.E., et al., *Expression of the Mf1 gene in developing mouse hearts: implication in the development of human congenital heart defects.* Dev Dyn, 1999. **216**(1): p. 16-27.

10. Mirzayans, F., et al., *Axenfeld-Rieger syndrome resulting from mutation of the FKHL7 gene on chromosome 6p25.* Eur J Hum Genet, 2000. **8**(1): p. 71-4.

11. Nishimura, D.Y., et al., *A spectrum of FOXC1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye.* Am J Hum Genet, 2001. **68**(2): p. 364-72.

12. Kawase, C., et al., *Screening for mutations of Axenfeld-Rieger syndrome caused by FOXC1 gene in Japanese patients.* J Glaucoma, 2001. **10**(6): p. 477-82.

13. Suzuki, T., et al., *A novel (Pro79Thr) mutation in the FKHL7 gene in a Japanese family with Axenfeld-Rieger syndrome.* Am J Ophthalmol, 2001. **132**(4): p. 572-5.

14. Panicker, S.G., et al., *Novel mutation in FOXC1 wing region causing Axenfeld-Rieger anomaly.* Invest Ophthalmol Vis Sci, 2002. **43**(12): p. 3613-6.

15. Saleem, R.A., et al., *Identification and analysis of a novel mutation in the FOXC1 forkhead domain.* Invest Ophthalmol Vis Sci, 2003. **44**(11): p. 4608-12.

16. Komatireddy, S., et al., *Mutation spectrum of FOXC1 and clinical genetic heterogeneity of Axenfeld-Rieger anomaly in India.* Mol Vis, 2003. **9**: p. 43-8.

17. Mortemousque, B., et al., *Axenfeld-Rieger anomaly: a novel mutation in the forkhead box C1 (FOXC1) gene in a 4-generation family.* Arch Ophthalmol, 2004. **122**(10): p. 1527-33.

18. Murphy, T.C., et al., *The wing 2 region of the FOXC1 forkhead domain is necessary for normal DNA-binding and transactivation functions.* Invest Ophthalmol Vis Sci, 2004. **45**(8): p. 2531-8.

19. Ito, Y.A., et al., *Analyses of a novel L130F missense mutation in FOXC1.* Arch Ophthalmol, 2007. **125**(1): p. 128-35.

20. Fuse, N., et al., *Novel mutations in the FOXC1 gene in Japanese patients with Axenfeld-Rieger syndrome.* Mol Vis, 2007. **13**: p. 1005-9.

21. Kulak, S.C., et al., *Mutation in the RIEG1 gene in patients with iridogoniodysgenesis syndrome.* Hum Mol Genet, 1998. **7**(7): p. 1113-7.

22. Alward, W.L., et al., *Autosomal dominant iris hypoplasia is caused by a mutation in the Rieger syndrome (RIEG/PITX2) gene.* Am J Ophthalmol, 1998. **125**(1): p. 98-100.

23. Doward, W., et al., *A mutation in the RIEG1 gene associated with Peters' anomaly.* J Med Genet, 1999. **36**(2): p. 152-5.

24. Perveen, R., et al., *Phenotypic variability and asymmetry of Rieger syndrome associated with PITX2 mutations.* Invest Ophthalmol Vis Sci, 2000. **41**(9): p. 2456-60.

25. Priston, M., et al., *Functional analyses of two newly identified PITX2 mutants reveal a novel molecular mechanism for Axenfeld-Rieger syndrome.* Hum Mol Genet, 2001. **10**(16): p. 1631-8.

26. Borges, A.S., et al., *Genetic analysis of PITX2 and FOXC1 in Rieger Syndrome patients from Brazil.* J Glaucoma, 2002. **11**(1): p. 51-6.

27. Phillips, J.C., *Four novel mutations in the PITX2 gene in patients with Axenfeld-Rieger syndrome.* Ophthalmic Res, 2002. **34**(5): p. 324-6.

28. Wang, Y., et al., *Novel identification of a four-base-pair deletion mutation in PITX2 in a Rieger syndrome family.* J Dent Res, 2003. **82**(12): p. 1008-12.

29. Brooks, B.P., et al., *A novel mutation in the PITX2 gene in a family with Axenfeld-Rieger syndrome.* Ophthalmic Genet, 2004. **25**(1): p. 57-62.

30. Lines, M.A., et al., *Characterization and prevalence of PITX2 microdeletions and mutations in Axenfeld-Rieger malformations.* Invest Ophthalmol Vis Sci, 2004. **45**(3): p. 828-33.

31. Saadi, I., et al., *An unusual class of PITX2 mutations in Axenfeld-Rieger syndrome.* Birth Defects Res A Clin Mol Teratol, 2006. **76**(3): p. 175-81.

32. Idrees, F., et al., *A novel homeobox mutation in the PITX2 gene in a family with Axenfeld-Rieger syndrome associated with brain, ocular, and dental phenotypes.* Am J Med Genet B Neuropsychiatr Genet, 2006. **141**(2): p. 184-91.

33. Vieira, V., et al., *Identification of four new PITX2 gene mutations in patients with Axenfeld-Rieger syndrome.* Mol Vis, 2006. **12**: p. 1448-60.

34. Saleem, R.A., et al., *Analyses of the effects that disease-causing missense mutations have on the structure and function of the winged-helix protein FOXC1.* Am J Hum Genet, 2001. **68**(3): p. 627-41.

35. Lehmann, O.J., et al., *Chromosomal duplication involving the forkhead transcription factor gene FOXC1 causes iris hypoplasia and glaucoma.* Am J Hum Genet, 2000. **67**(5): p. 1129-35.

36. Lehmann, O.J., et al., *Ocular developmental abnormalities and glaucoma associated with interstitial 6p25 duplications and deletions.* Invest Ophthalmol Vis Sci, 2002. **43**(6): p. 1843-9.

37. Ekong, R., et al., *Chromosomal anomalies on 6p25 in iris hypoplasia and Axenfeld-Rieger syndrome patients defined on a purpose-built genomic microarray.* Hum Mutat, 2004. **24**(1): p. 76-85.

Chapter 4:

General Discussion

In the preceding chapters, I have presented my own understanding, work and the contribution my work will have towards the molecular understanding of the Axenfeld-Reiger Syndrome. Here, I would like to discuss the benefits that are incurred from extensive characterization of CNVs and some future directions that this study could take.

Importance of genomic rearrangements' characterization

The molecular mechanisms of genomic rearrangements remained incompletely defined as the majority of reported genomic rearrangements are not characterized at the base pair level. The two most widely accepted mechanisms of genomic rearrangements are NAHR and NHEJ. However, NAHR and NHEJ were proposed prior to the advent of high-resolution array CGH. As such, these mechanisms were proposed based on the characterization of relatively large CNVs that were primarily identified with locus specific fluorescent in situ hybridization (FISH), chromosome painting, telomeric FISH, etc. With the advent of high resolution array CGH, submicroscopic CNVs have been shown to be exceedingly common [1-3]. Thus it remained to be determined if NAHR and NHEJ could truly account for the genesis of these recently appreciated but substantially large numbers of submicroscopic CNVs.

A key state of uncovering the precise mechanisms of genomic rearrangements, and in turn the genesis of CNVs, is to narrow the strand exchange interval. This can be achieved by amplifying the rearrangements specific junction fragment and then subsequently sequencing it to map the breakpoints at the nucleotide level. Such an approach, as followed in this thesis and in other studies, has proved invaluable in providing insights into the molecular mechanisms of genomic rearrangements.

Prior to this study, the benefits of genomic rearrangements' characterization were highlighted by many others. For instance, the preferred substrate for NAHR has long been found to be LCRs, and to some extent Alu elements [4-7]. However, it was the detection of rearrangement-specific junction fragments with pulsed-field gel electrophoresis (PFGE) and subsequent sequencing studies in both Charcot-Marie-Tooth disease type 1A (CMT1A) and the hereditary neuropathy with liability to pressure palsies (HNPP), that revealed a 557bp "recombination hotspot" within 24kb LCRs (CMT1A-Rep) [8-10]. Such observations, leading to the hypothesis that *cis*-acting sequences can stimulate increased potential for double strand breaks during strand exchange between LCRs, provided insights into the NAHR-mediated rearrangements.

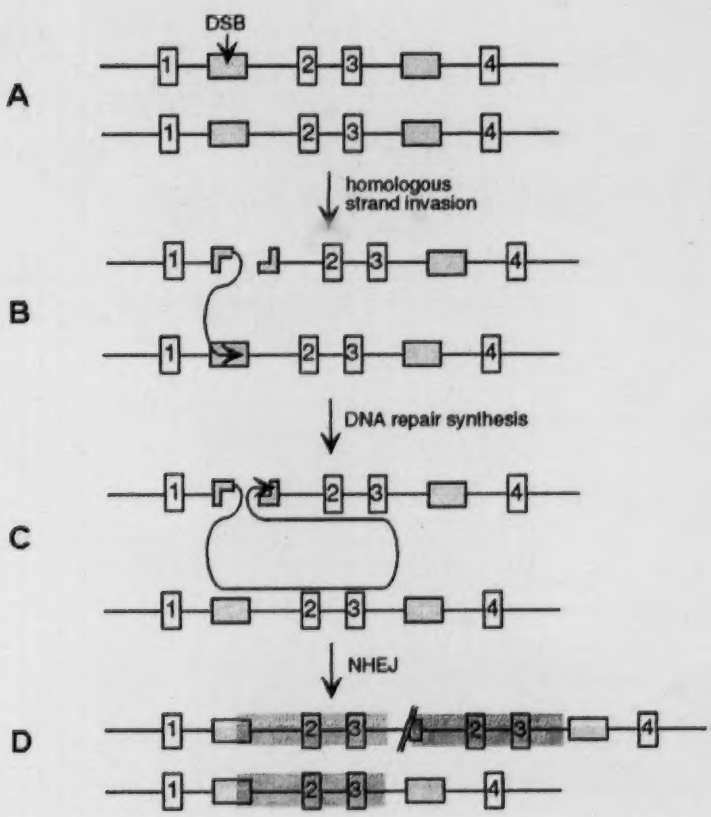
The benefits of genomic rearrangements' characterization were further highlighted in Woodward *et al.*[11], and very recently in Lee *et al.*[12]. Characterization of

Xq22 rearrangements (segmental duplications), encompassing the proteolipid protein 1 gene (*PLP1*), in only 22% (13 out of 59) of PMD cases revealed that a complex mechanism-coupled NAHR and NHEJ-underlies these genomic rearrangements (Figure 4.1). In 22% or 13 PMD cases, Woodward *et al.* observed normal reference genomic sequence at one end of the duplicated region – suggestive of an error-free homologous repair process, and insertion or micro homology of 1-6 nucleotides at the other end- suggestive of a non-homologous repair process. Such observation extended the range of mechanisms associated with genomic rearrangements. Very recently, Lee *et al.* in 2008, with combined high-resolution oligonucleotide array CGH and breakpoint sequence analysis of nonrecurrent duplications of PMD, proposed a new replication-based mechanism – the Fork Stalling and Template Switching (FoSTes) - for genomic rearrangements.

Presently, my work, characterization of 6p25 CNVs at the base pair level, reveals a novel mechanistic spectrum that underlies the genesis of CNVs. Five unique genomic rearrangements, three segmental duplications [Figures 2.2 (pedigrees # 1-2 and #3-6) and 2.3 (pedigree #7)] and two segmental deletions [Figures 2.4 (pedigree #8) and 2.5 (pedigree #9)], were characterized at the base pair level which revealed their non-recurrent nature. The characterization also revealed an unprecedented size spectrum (~30 – 1220 kb) and provided evidence for the use of small repetitive sequences (Alu, simple repeats etc.), rather than LCRs, in mediating genomic rearrangements.

Figure 4.1 Coupled NAHR and NHEJ.

(**A-D**) A proposed mechanism of coupled NAHR and NHEJ, taken from [13].
A. Double strand break, at a homologous region, could be the initiating event of NAHR.
B. Following DSB, DNA repair synthesis results in homologous strand invasion and DNA sequence of the invaded template is copied.
C. The DNA repair is complete by NHEJ.
D. At the end of this process duplication of a genomic region may occur. The duplicated region is pink shaded while the normal genomic region is green shaded.

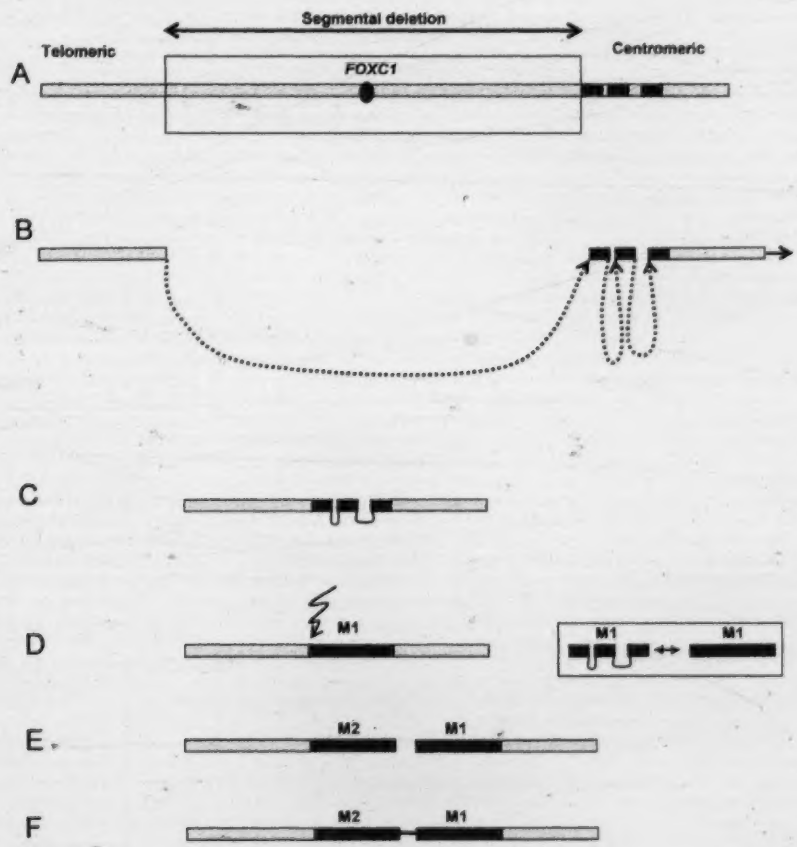


Such observations extend the range of substrate associated with genomic rearrangements. In most of the characterized 6p25 CNVs, the use of differing degrees of two mechanisms - NAHR and NHEJ- were observed (Figure 2.7). The coupled NAHR-NHEJ mechanism, originally proposed to maintain genomic integrity in mammalian cells [13], has only been suggested to mediate the *PLP1*-encompassing duplication (Xq28) in PMD [11]. However, genomic rearrangements using a spectrum of NAHR and NHEJ are unprecedented. Also, the coupled NAHR-NHEJ mechanisms has been proposed to mediate segmental duplications (Xq28) only [11], while in this study a spectrum of NAHR and NHEJ in *both* segmental duplications and deletions has been observed.

One of the most important and interesting parts of this thesis is the characterization of the genomic rearrangement in pedigree #8 (Figure 2.4). A very complex mechanism was found to be associated with this genomic rearrangement. Although this complex mechanism is similar to the proposed FoSTeS [12], we observed several unique features in the genomic architecture of pedigree #8 which could not be explained by FoSTeS alone (Figure 2.4). Firstly, in FoSTeS the lagging strand is proposed to switch template following fork stalling. However, the genomic architecture in pedigree #8 is better explained when the replication fork is hypothesized to jump sequence on encountering impediments (Figure 4.2). Secondly, the high order of symmetry observed between the motifs (M1 and M2)

Figure 4.2 Probable mechanism for the origin of genomic architecture in pedigree #8

A. Schematic representation of the genomic region encompassing *FOXC1* (red circle). The genomic region experiencing the segmental deletion is boxed.
B. The movement of the replication fork is indicated by dotted arrows. The regions of centromeric sequence that contributed to the origin of the motif M1 are depicted with black and brown rectangles.
C. Replication error leads to segmental deletion and the origin of motif M1.
D. Double strand break occurs at the telomeric end of the motif M1.
E. DNA repair process generates the motif M2 from M1.
F. M1 and M2 are joined during (solid blue line) the completion of the DNA repair process.

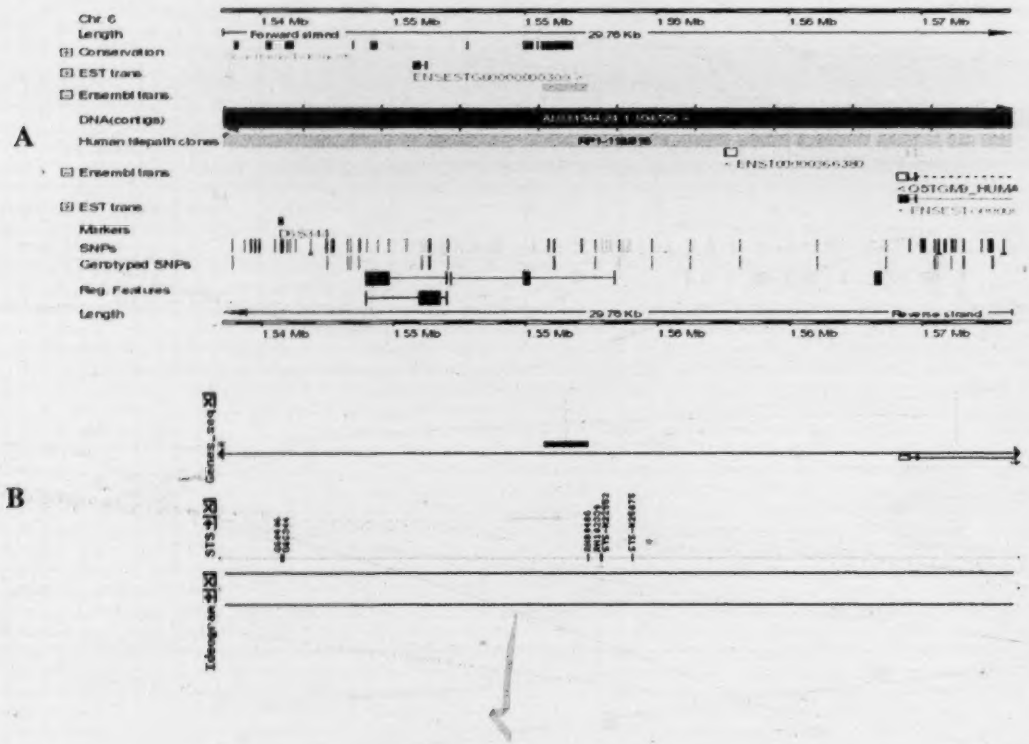


is suggestive of the fact that one arises from the other. The most parsimonious explanation for this is the occurrence of DSB followed by erroneous DNA repair. Again, such complexities of mechanisms are not known to be associated with FoSTeS. Finally, FoSTeS is described in an X-linked disorder (PMD) and has been shown to mediate only duplications, while we observed a near to similar mechanism in a segmental deletion associated with an autosome (chromosome 6)-linked ocular disorder. These observations extend the range of genomic rearrangements and disorders that can be explained by replication-based mechanisms.

In contrast to the spectrum of NAHR - NHEJ and replication based mechanisms, this thesis also provides evidence of only NHEJ in mediating genomic rearrangements. In pedigree # 9, characterization of a small (~30kb) 6p25 deletion revealed NHEJ to be the sole underlying mechanism of genomic rearrangement (Figure 2.5). Interestingly, the deletion rearrangement specific junction fragment was only present in two ARS-affected siblings while it was absent in the phenotypically normal parents (Figure 2.6D). It is highly unlikely that this 6p25 deletion arose *de novo* and separately, since the genomic architecture of this deletion is identical at the base pair level in these two siblings. These evidences suggest that the deletion arose from one of the parents who is somatically mosaic. The parental origin of this 6p25 CNV awaits microsatellite marker genotyping. The probable SNPs and microsatellite markers that can be used for genotyping are illustrated in Figure 4.3.

Figure 4.3: Montage illustrating the probable microsatellite markers and SNPs that can be used for haplotype analysis in pedigree#9.

A. View of the deleted region in pedigree #9 from Ensembl.
B. View of the deleted region in pedigree #9 from NCBI.



Extensive characterization of 6p25 genomic rearrangements not only revealed the underlying mechanistic spectrum of these genomic rearrangements but also identified a discrepancy between array CGH predicted and actual breakpoints of a duplication rearrangement (Figure 2.3). In another separate incidence, the array CGH result differed from the microsatellite marker analysis (Figure 3.1). These observations are compatible with the fact that array CGH, though a very powerful technique in identifying CNVs, is not infallible. The high resolution of array CGH makes it a preferred choice of CNV identifying technology which is evident from the fact that the majority of the genome wide studies for estimating the prevalences of CNVs have used this technology in modified forms [2, 3, 14-20]. Based on such capacity of array CGH in identifying CNVs, the clinical implementation of this technology has also been suggested [21-23]. However, the findings of this thesis and different caveats of array CGH (Chapter 1- Array CGH), strongly suggest that array CGH results should be validated with alternative techniques, prior to its clinical implementation.

In summary, the study presented in this thesis adds complexities and provides novel insights into the mechanisms of 6p25 genomic rearrangements, and by implication ARS. The observation of a novel spectrum of recombination, DNA repair and replication based mechanism is unprecedented. The evidence of small repetitive sequences mediating genomic rearrangements broadens the range of substrate that

can mediate genomic rearrangements and thus has implications for the genesis of copy number variations in other regions of the human genome. This thesis highlights the benefits of undertaking extensive studies necessary to characterize CNVs at the base pair level.

Future Directions

Towards identifying novel genetic factors in ARS

One of the key factors for identifying novel genetic factors of a particular phenotype is the availability of a large panel of patients with that particular phenotype, as in this case-ARS. We have the largest collection of ARS pedigrees worldwide (total affected: 148), which provides an excellent platform to embark upon the goal of identifying novel genetic factors of ARS. Furthermore, availability of cryopreserved leukocytes from ARS cases ensures high quality DNA-essential for array CGH.

Preliminary progress has been made towards the above goal by screening *FOXC1* and *PITX2* for mutations and/or CNVs in the ARS panel, however, much remain to be done. The *FOXC1* and *PITX2* mutation or CNV negative ARS samples will be further subjected to region specific or whole genome oligonucleotide array CGH to identify putative CNVs. This criterion needs to be carefully determined. Both approaches have their own caveats resulting from the fact that the number of probes

to play with *per array* remains constant - max probe density per array in HD2 platform (NimbleGen®) is 2.1 million probes. Since the probe density is limited, the genome wide array CGH will compromise the resolution at which the CNVs will be detected. However, this is an unbiased approach that aims to identify CNVs throughout the genome. On the other hand, customized array CGH will focus on the areas of interest and thus represents a biased approach. However, the latter approach allows high-resolution at which CNVs can be identified. As such, the challenge is to strike a balance between the extent of coverage and the resolution of array CGH.

The latest HD2 array from NimbleGen® is divisible into 12 subarrays, each comprising ~ 130,000 probes. These subarrays can be further customized to achieve desired resolution across a given genomic interval. After the hybridization, the challenge will be to distinguish benign CNVs from pathogenic ones. Comparing the putative CNVs with the available CNV databases [19, 24] will serve as the initial guide. However, the CNV databases are far from being perfect due to different associated caveats, as described in Chapter 1. To further facilitate the process of identifying the pathogenic CNVs, intra-familial controls (unaffected relatives of ARS cases) and/or ethnically matched controls (same ethnicity as that of the ARS cases but without any ocular abnormalities) will be screened for putative CNVs. The putative CNVs found *both* in normal and ARS cases may not be considered for future studies. The causal association of putative CNVs with ARS will be confirmed

by amplifying breakpoints spanning junction fragment, primarily, in ARS affected individuals. Such an extensive approach is expected to identify novel genetic factors of ARS.

Simple molecular diagnostic test of ARS/developmental glaucoma

Current estimate predicts that ~80 million people will have glaucoma by 2020 [25]. Early diagnosis is crucial to limit the progression of glaucoma. However, even in developed western countries the portion of undiagnosed glaucoma cases is estimated to be ~50% [26]. There exist several techniques that can be employed to detect some of the typical manifestations of glaucoma- such as elevated IOP [measured by tonometry], visual field loss [measured by perimetry] and anatomical changes in the optic nerve [measured by ophthalmoscopy, scanning laser polarimetry (GDx), confocal laser ophthalmoscopy or Heidelberg Retinal Tomography (HRT II) and optical coherence tomography (OCT)]. However, these typical manifestations are not found in all glaucoma cases. For instance, normal tension glaucoma (NTG) is not characterized by elevated IOP. Coupled with this, early glaucoma patients can lose ~50% of their ganglion cells before exhibiting visual field defects [27]. All these factors make the clinical diagnosis of glaucoma challenging. As such, a diagnostic method, independent of these typical glaucoma manifestations, is highly desirable for early detection of patients with or at high risk of developing glaucoma.

In this study, three junction fragments, indicative of three specific 6p25 CNVs, were found in 148 patients with ARS or glaucoma phenotype. The first three primer pairs in Table 2.2 have been used to amplify these junction fragments by PCR. It is worth mentioning here that genomic rearrangements in most of the ARS patients were previously identified through FISH and/or microsatellite marker genotyping or real time PCR, prior to the PCR-based amplification of rearrangements' specific junction fragments. However, considering the rarity of ARS and the fact that simple PCR can amplify ARS-causing genomic rearrangements-specific junction fragments in 148 ARS patients, suggest that a PCR-based molecular test could prove beneficial in identifying the genetic cause of a significant portion of ARS cases. In addition, the PCR based molecular test could also be employed for prenatal screening of ARS, especially in families with known genetic history of ARS or glaucoma. Recently, genotype-phenotype correlation study in ARS patients has revealed that patients with *FOXC1* duplications have more severe prognosis in glaucoma development than patients with *FOXC1* mutations [28]. Such data further highlight the importance of deciphering the genetic cause of ARS for better prognosis and in turn, patients' treatment.

Probably the greatest limitation of the PCR-based molecular test is that negative results are elusive. The PCR based molecular test, currently, can only identify six 6p25 rearrangements that lead to ARS, leaving other ARS-associated 6p25 rearrangements unidentifiable. Further molecular tests will be required to confirm if

PCR-negative ARS patients have genomic alterations. In addition, the possibility of a PCR-based test of ARS is surrounded by many unanswered questions: How successful will the clinical implementation of the test be? What is the true portion of ARS/developmental glaucoma that can be detected with this test? Should this test be implemented for screening individuals with genetic history of ARS/glaucoma? And above all – When can this test be found in the routine molecular diagnosis of such ocular disorders? Only the future holds the answer to these questions, but following the beliefs of Einstein - "*I never think of the future-it comes soon enough*"- we know that the answers are soon to follow.

Bibliography

1. Carter, N.P., *Methods and strategies for analyzing copy number variation using DNA microarrays.* Nat Genet, 2007. **39**(7 Suppl): p. S16-21.

2. Sharp, A.J., Z. Cheng, and E.E. Eichler, *Structural variation of the human genome.* Annu Rev Genomics Hum Genet, 2006. **7**: p. 407-42.

3. Tuzun, E., et al., *Fine-scale structural variation of the human genome.* Nat Genet, 2005. **37**(7): p. 727-32.

4. Kanaar, R., J.H. Hoeijmakers, and D.C. van Gent, *Molecular mechanisms of DNA double strand break repair.* Trends Cell Biol, 1998. **8**(12): p. 483-9.

5. Stankiewicz, P. and J.R. Lupski, *Genome architecture, rearrangements and genomic disorders.* Trends Genet, 2002. **18**(2): p. 74-82.

6. Shaw, C.J. and J.R. Lupski, *Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease.* Hum Mol Genet, 2004. **13 Spec No 1**: p. R57-64.

7. Lupski, J.R. and P. Stankiewicz, *Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes.* PLoS Genet, 2005. **1**(6): p. e49.

8. Lopes, J., et al., *Homologous DNA exchanges in humans can be explained by the yeast double-strand break repair model: a study of 17p11.2 rearrangements associated with CMT1A and HNPP.* Hum Mol Genet, 1999. **8**(12): p. 2285-92.

9. Reiter, L.T., et al., *Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients.* Am J Hum Genet, 1998. **62**(5): p. 1023-33.

10. Reiter, L.T., et al., *A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element.* Nat Genet, 1996. **12**(3): p. 288-97.

11. Woodward, K.J., et al., *Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination.* Am J Hum Genet, 2005. **77**(6): p. 966-87.

12. Lee, J.A., C.M. Carvalho, and J.R. Lupski, *A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders.* Cell, 2007. **131**(7): p. 1235-47.

13. Richardson, C. and M. Jasin, *Coupled homologous and nonhomologous repair of a double-strand break preserves genomic integrity in mammalian cells.* Mol Cell Biol, 2000. **20**(23): p. 9068-75.

14. Freeman, J.L., et al., *Copy number variation: new insights in genome diversity.* Genome Res, 2006. **16**(8): p. 949-61.

15. Khaja, R., et al., *Genome assembly comparison identifies structural variants in the human genome.* Nat Genet, 2006. **38**(12): p. 1413-8.

16. Locke, D.P., et al., *Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome.* Am J Hum Genet, 2006. **79**(2): p. 275-90.

17. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

18. Wong, K.K., et al., *A comprehensive analysis of common copy-number variations in the human genome.* Am J Hum Genet, 2007. **80**(1): p. 91-104.

19. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome.* Nat Genet, 2004. **36**(9): p. 949-51.

20. Sebat, J., et al., *Large-scale copy number polymorphism in the human genome.* Science, 2004. **305**(5683): p. 525-8.

21. Itoh, F., H. Yamamoto, and K. Imai, *[Clinical application of DNA array system in human gastrointestinal cancer].* Hokkaido Igaku Zasshi, 2002. **77**(2): p. 143-4.

22. Takahashi, Y., et al., *Clinical application of oligonucleotide probe array for full-length gene sequencing of TP53 in colon cancer.* Oncology, 2003. **64**(1): p. 54-60.

23. Wa, C.V., et al., *Clinical application of array-based comparative genomic hybridization to define the relationship between multiple synchronous tumors.* Mod Pathol, 2005. **18**(4): p. 591-7.

24. Zhang, J., et al., *Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome.* Cytogenet Genome Res, 2006. **115**(3-4): p. 205-14.

25. Quigley, H.A. and A.T. Broman, *The number of people with glaucoma worldwide in 2010 and 2020.* Br J Ophthalmol, 2006. **90**(3): p. 262-7.

26. Quigley, H.A. and S. Vitale, *Models of open-angle glaucoma prevalence and incidence in the United States.* Invest Ophthalmol Vis Sci, 1997. **38**(1): p. 83-91.

27. Lehmann, O.J., *The Role of forkhead Genes in Eye Development and Disease*, in *Molecular genetics*. 2002, University College London: London.

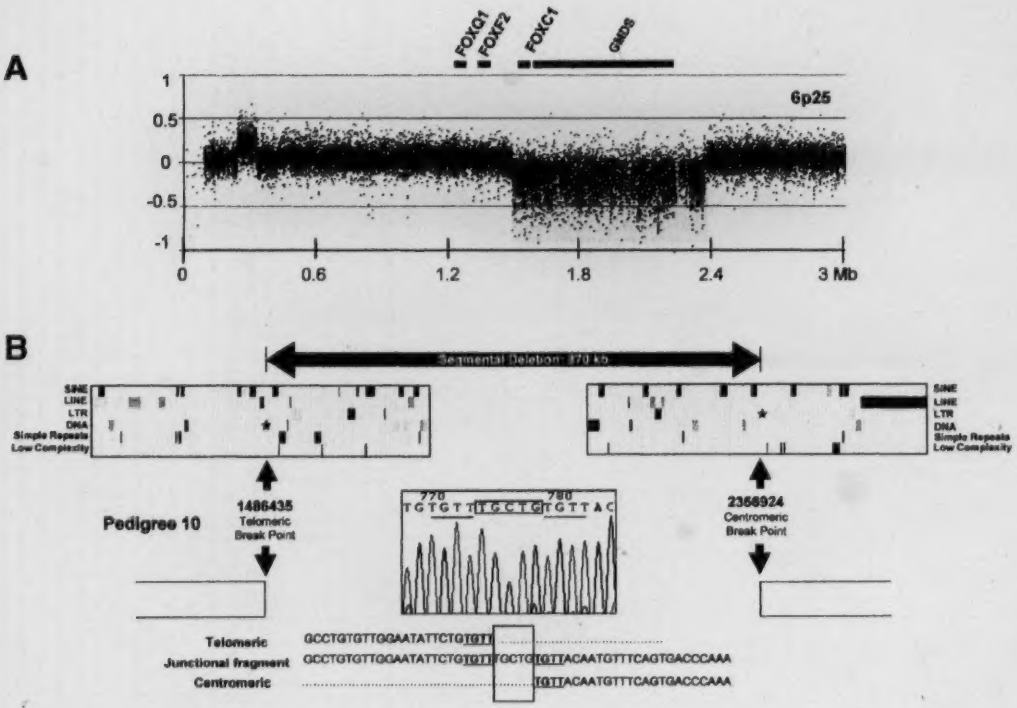
28. Strungaru, M.H., I. Dinu, and M.A. Walter, *Genotype-phenotype correlations in Axenfeld-Rieger malformation and glaucoma patients with FOXC1 and PITX2 mutations.* Invest Ophthalmol Vis Sci, 2007. **48**(1): p. 228-37.

Appendix A.

Montage illustrating array CGH result and genomic architecture of a 6p25 segmental deletion.

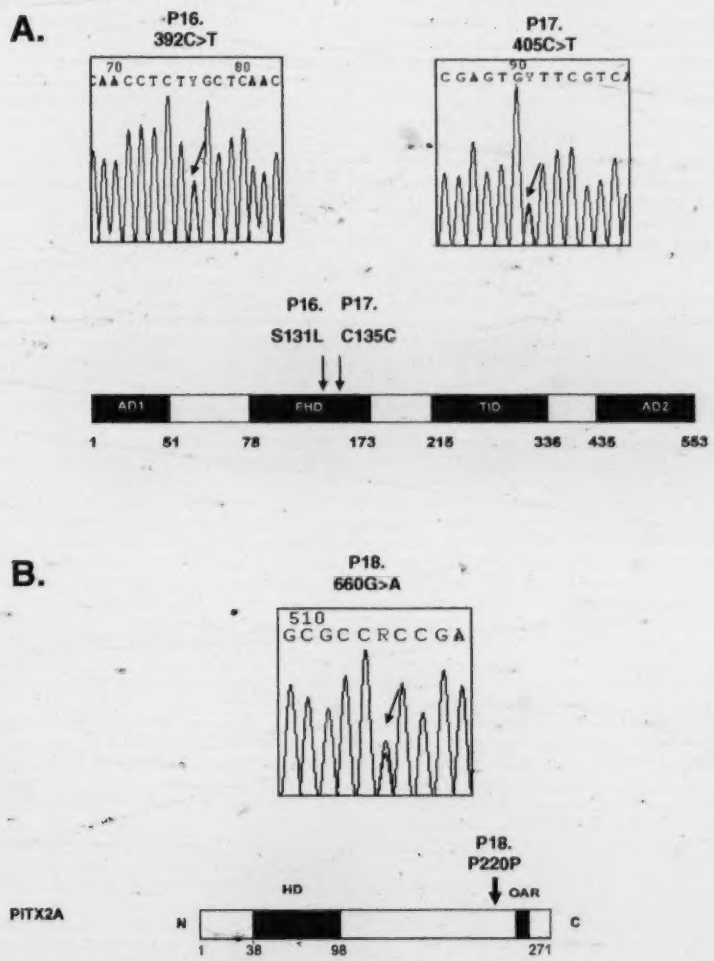
A. The copy number variant region is shaded light blue. The genes encompassed by the CNV are depicted in brown. The relative location of the FOX triplet (*FOXFQ1*, *FOXF2*, *FOXC1*) is also denoted.

B. The genomic architecture of the 6p25 segmental deletion. The telomeric and centromeric breakpoints are indicated with stars and the extent of the deletion (~870kb) is denoted by red filled arrow. The chromatogram displays the insertion of 6bp nucleotides (boxed) and the microhomology of 4bp nucleotides (underlined) at the rearrangement's junction.



Appendix B.

FOXC1 and *PITX2* sequence alterations identified in ARS cases.

A. Two heterozygous mutations identified in *FOXC1*. The alteration 392C>T result in S131L, whilst the alteration 405C>T produces a silent change (C135C). Both alterations lie in the forkhead domain of *FOXC1*.

B. One silent change P220P, corresponding to 660G>A, was identified in *PITX2*.